DOCUMENT RESUME

ED 234 636 FL 013 958

AUTHOR Bordan, Robert; Garcia, Maryellen
TITLE Discourse-Sensitive Measurement of Language Development in Bilingual Children.
INSTITUTION National Center for Bilingual Research, Los Alamitos, Calif.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO NCBR-07-82
PUB DATE Jul 82
NOTE 60p.
AVAILABLE FROM National Center for Bilingual Research, 4665 Lampson Ave., Los Alamitos, CA 90720.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Bilingualism; Child Language; Children; *Discourse Analysis; English; *Evaluation Methods; Language Acquisition; *Language Proficiency; Spanish; Speech Communication
IDENTIFIERS Mean Length of Utterance

ABSTRACT

The use of observation of natural language interaction as a measure of language proficiency and the impact of discourse characteristics on children's use of Spanish and English as measured by length of utterances are examined. The goal of this observational approach to measuring language proficiency is to distinguish between the effects of change in discourse contexts and change in the language proficiency of individuals over time. Maximum likelihood techniques are used to estimate the effects of discourse contexts on length of utterance, and the probability that utterances will be as long as those observed in each discourse context is then calculated. This probability becomes the basis for constructing a weighted index of utterance length. This approach was tested on language samples obtained from Spanish/English bilingual children between the ages of 4 and 10 and compared to other indicators of language ability. It was found that discourse function or context appears to influence utterance length. (Author/RW)

ED234636

DISCOURSE-SENSITIVE MEASUREMENT OF LANGUAGE DEVELOPMENT IN BILINGUAL CHILDREN

Robert Berdan and Maryellen Garcia

July, 1982

07-82

FL013958

national center for bilingual research, 4665 lampson avenue, los alamitos, CA 90720, telephone (213) 598-0481

The material in this publication was prepared under Cooperative Agreement 00-CA-80-0001 from the National Institute of Education, Department of Education. Its contents do not necessarily reflect the views of the National Institute of Education or of any other agency of the United States government, and no official endorsement should be inferred.

ABSTRACT

In this paper we explore several problems in the use of natural language interaction as a measure of language proficiency. The approach developed here is based on observation of natural language interaction, with a goal of distinguishing between the effects of change in discourse contexts and change in the language abilities of individuals over time. We use maximum likelihood techniques to estimate the effects of discourse contexts on length of utterance. We then calculate the probability that utterances will be as long as those observed in each discourse context. This probability becomes the basis for constructing a weighted index of utterance length. Our approach is tested on language samples from Spanish/English bilingual children and compared to other indicators of language ability.

DISCOURSE-SENSITIVE MEASUREMENT OF LANGUAGE DEVELOPMENT IN BILINGUAL CHILDREN*

Robert Berdan and Maryellen Garcia

Introduction

In this paper we look at the impact of a variety of discourse characteristics on children's use of Spanish and English, as measured by length of utterance. Discourse contexts intervene strongly in the relationship between length of utterance and language development. Procedures are introduced here for distinguishing between the effects of discourse context, and changes in the language abilities of individuals over time. This allows the possibility of more accurately measuring growth in language proficiency by observation of natural language interaction.

The present work is part of on-going longitudinal studies of language development in bilingual contexts (Garcia, Veyna-Lopez, Siguenza & Torres, 1982). These studies include primarily naturalistic observations of children in Spanish-English and Korean-English environments on a monthly schedule. Over the course of the study, various of the children will range in age from approximately four years to ten years. The children are being observed in their use of English, as well as in the language of the home, Spanish or Korean. The study is being undertaken to document the nature of the language development process for children in bilingual contexts, with particular interest in relating that process to educational practice.

Desired Characteristics in a Measure of Language Development

These characteristics of the longitudinal study impose a series of constraints or desired properties for an acceptable indicator of language development. The need, however, is not unique to this

*A number of people have been extremely helpful in the development of this paper. Consuelo Siguenza did the discourse coding on which the analysis is based. Dr. Pascale Rousseau reviewed an earlier draft, and provided many helpful comments, particularly on the use and interpretation of the maximum likelihood estimates. Dr. Alvin So performed the many passes of computer data analysis.

longitudinal study. Both in research and in practical applications, the need for improved measures of language development continues. We identify the following five characteristics as particularly desirable:

1. It should be fine-grained, able to detect change over fairly small units of time, i.e., months or aggregates of several months.

2. It should be language-independent, or capable of comparable forms across languages of fundamentally differing linguistic structure.

3. It should be continuous across the age range 4-10, and desirably be continuous from infancy to adulthood.

4. It should be capable of measuring language in use, directly or indirectly, rather than the abstract notion of "knowledge of language."

5. It should be sensitive to a school-based notion of language as proficient communication, rather than narrowly based on language as linguistic structure.

This is a highly idealized list of properties. No such measure now exists; it may not be obtainable. Nonetheless, the analytic approach taken below seems to offer at least some promise in each of these areas.

Measures Derived from Natural Discourse

Measures of language development can be grouped in a number of different ways. Among these we find it useful to distinguish between those which employ primary characteristics of language, and those which employ secondary, or derived characteristics. By primary characteristics we refer to measures which derive from linguistically well-defined elements and relationships. Indicators of the use of a great variety of language characteristics fall into this category, including such things as inflectional morphemes (e.g., plural markers, verb agreement markers), relative clauses, and various other syntactic markers or constructions.

All of these elements of language do develop over time. Charting change in their frequency and distribution through time is an important characterization of the language development process. Many of these indicators, however, tend to stabilize early in the acquisition process. Research has shown that for monolinguals they tend to be largely acquired by the onset of schooling, or shortly thereafter (cf. Berko-Gleason, 1971; Menyuk, 1971). Those indicators which are most important for distinguishing specific stages of language development in children over a wide age range seem to occur quite infrequently in natural discourse, thus requiring specifically structured elicitation procedures to occasion reliable frequencies of observation. Such structuring is not possible in a study that focuses primarily on the child's natural interaction with a variety of interlocutors.

Secondary indicators, on the other hand, measure language development in global, rather than particularistic, terms. Some such indicators that have been suggested for educational applications are based on T-units (Hunt, 1965) or communication units (Loban, 1976), and overall measures of utterance length (Brown, 1973; Cazden, 1968). These secondary measures can be distinguished from primary indicators in the following sense: Language learners can be said to be acquiring the use of plural markers or agreement markers, or any other primary language characteristics, with increasing probability across time, and this is readily explained in terms of theoretically well-motivated linguistic processes.

On the other hand, to say that children are acquiring the use of more words per sentence is not well defined in any generally supported theory of language or cognitive development. This is not to say that children do not use longer sentences as their language ability develops. They do; but children also engage in increasingly complex topics of discourse, with increased demands for information transfer. Children's control of primary language characteristics, such as the processes of syntactic embedding, also increase. Thus, children have at their disposal an increasing array of linguistic devices for expressing increasingly complex messages. These primary processes underlie the derived or secondary relationships between number of words and the syntactic units containing them. Analysis of the primary linguistic and cognitive characteristics, however, is extremely complex and costly. In some cases there is not even a well agreed upon basis for the classification of observations. The secondary characteristics, however, expressed as a ratio of lexical units to syntactic units, confound language and cognitive and social development in a way that may frustrate academic researchers, but which seems to characterize fairly reasonably the somewhat confounded notion of language proficiency which educators find most appealing. As outlined below, we have used in our analyses the secondary measure of words per sentence, or utterance length, as an indicator of development.

Mean length of utterance: Measuring discourse effects. Mean length of utterance (MLU) has been widely used to report early stages of acquisition both for English-speaking (Brown, 1973; Bloom, 1970; Cazden, 1968) and Spanish-speaking children (Brisk, 1972; Peronard, 1977; Padilla & Liebman, 1982). Despite the fact that large studies show that MLU increases monotonically with age through the school years (at least for samples of writing; see, for example, Hunt, 1965; O'Donnell, Griffin, & Norris, 1967), MLU in any of its variously calculated forms has not shown much usefulness past the two or three word stage of development. A variety of problems related to clinical use of MLU as an indicator of development are treated in a collection of articles reprinted in Longhurst (1974). Many of these problems seem to derive from the many ways in which discourse structure influences utterance length, and the confounding that is introduced by sampling fluctuation across different discourse contexts.

1. Ellipsis. One of the most obvious of these is the phenomenon of ellipsis. The rules for ellipsis in English discourse are rather complex (cf. Halliday & Hassan, 1976). Ellipsis may be characterized generally, however, as the nonrepetition of identical information across adjacent turns in a conversation. Consider the following exchange:

Interviewer: What do you think they're going to do after they finish eating?

Child: Play. (full ellipsis)

This response contains just one word. Yet it is fully as appropriate for the discourse as would any of the following forms have been:

> I think they're going to play after they finish eating. (no ellipsis)
>
> I think they're going to play. (partial ellipsis)
>
> After they finish, they're going to play. (partial ellipsis)

The more extended responses might have offered positive evidence of fairly well developed English fluency; the use of the most elliptic form, however, offers no negative evidence. In some instances the failure to use ellipsis where it is possible makes the conversation sound unnatural.

The discourse appropriate ellipsis of the exchange above is quite different from the child's turns below, which are the same length and are superficially similar, but do not show syntactically well-formed ellipsis:

Interviewer: What do you do when you play with your friends?

Child: Game.

Interviewer: What kind of games do you play?

Child: Toy.

Interviewer: Okay, do you have a favorite toy? . . . Tell me about your favorite toy, okay?

Child: Space.

Interviewer: It's a space toy? What do you do with it?

Child: Playing.

These turns by the child are "contingent" on the prior turn, in the sense developed by Bloom and colleagues (e.g., Bloom, Rocissano & Hood, 1976). They share the same topic as the previous turn, and add new information to it. They are not elliptic in the conventional sense, but are rather what Braine (1974:456) termed "pseudo-elliptical." They cannot be said to result from a discourse reduction process, but are instead essentially holophrastic.

In summary, elliptic utterances may be considerably shorter than their nonelliptic paraphrases, but offer no evidence for less well developed language proficiency. Conversely, some very short elliptic utterances indicate ability to comprehend and use dialogue appropriately, thus showing greater language proficiency than some equally short, but non-elliptic utterances. Clearly, this phenomenon confounds the use of utterance length as a measure of language development. Any measure based on length must either control the relative frequency of ellipsis in the sample of language analyzed, or it must provide a means of weighting utterances of the same length differently in order to compensate for ellipsis effects.

2. <u>Discourse function</u>. Other aspects of discourse structure and function have similar confounding effects. The function of a particular turn, or the function of the immediately prior turn in the discourse both influence what information is exchanged, and thus also influence the length and complexity of the utterances. This is apparent in the following exchange:

Interviewer: Do you like science?

Child: Yes. I like it.

Interviewer: What else do you learn in your science class?

Child: About water.

Interviewer: What about water?

Child: Water becomes a liquid and the gas, no the air, becomes water.

The interviewer's first question is a request for a specific piece of information. It is a yes/no question; it can be answered appropriately with a single word, or, as in this case, with a pronominalized rephrasing of the question. The interviewer's second question is another request for information, referring to a prior topic in the conversation. It is readily answered with a simple noun phrase. The third question is a request for elaboration on the child's immediately preceding turn. The child's response could vary widely in the amount of elaboration provided, but almost all of the alternatives would require a full sentence as a response. The request for elaboration is also an invitation for the child to provide an extended response on a topic introduced by the child. Utterances in such a discourse context tend to be rather longer than responses to requests for specific information.

Attempts to resolve discourse effects. Some researchers, frustrated by the instability of MLU as a measure of language development, have attempted to control the discourse context in which language is sampled. This is frequently done by supplying a context, often pictures, and by specifying a task, usually description or story telling. What the adult participant may or may not do or say is also controlled. These are discourse variables of a higher order than what we have considered here. Observed differences across these macro discourse variables may well turn out to be due to differing rates of occurrence of the more micro discourse variables that are considered here. In any event, such things as ellipsis potential and function of utterance may be difficult and in some cases impossible to control in any naturalistic language use or elicitation situation.

Considered jointly, these three discourse characteristics--ellipsis, function of previous utterance, function of utterance--seem to differentiate utterance length to a greater extent than does chronological age or length of exposure to language in the sample of children in the longitudinal studies. The analyses presented here provide a way of incorporating this and other discourse-related information into the measurement of utterance length. This reduces the sampling problem and tends to stabilize any length measure based on naturalistic conversation.

Methodology

The findings reported here result from continued analysis of the data generated by NCBR's Longitudinal Studies of Language Development in Bilingual Contexts. Some aspects of the particular data sample used here were reported in Berdan & Garcia (1982).

Participants. The sixteen children in this data set ranged in age from 3;8 to 9;8 at the time the language samples were collected in the Summer of 1981. The children are eight sibling pairs from throughout the greater Los Angeles area. The home language for these children ranges from almost exclusively Spanish to largely English. The children are in a variety of regular and special instructional programs in their schools. Characteristics of the children, their homes and their schools are detailed in Garcia et al. (1982). For the longitudinal study, these children are visited by a bilingual fieldworker monthly in their homes and in their schools. Each session is tape recorded.

Elicitation procedures. The data for this analysis were elicited using the picture description and story telling task of the Basic Inventory of Natural Language (BINL) (Herbert, 1979). Following the general procedures for the administration of the BINL, the children were asked to describe various pictures which they selected from a set of culturally diverse color pictures. The sessions were conducted in the children's homes; both focal siblings were present throughout, and were encouraged to interact during the course of the session. In some cases there were also other family members or neighborhood children present. The sessions were tape recorded and subsequently transcribed. The fieldworker directed that for each child the task was to be done first in English and then in Spanish. The goal of the session was to elicit at least 50 utterances in English and in Spanish from each child. In some cases this was not possible, particularly in the weaker language of the younger children. Three of the children did not produce even ten utterances in English, and were excluded from the analysis of the English data.

Coding. Utterances were extracted from the transcripts in each language in accordance with BINL procedures specified in Herbert (1979). The lists of extracted utterances were then sent to the BINL publisher for machine scoring. In that process the utterances were edited to exclude sentence partials that do not conform to the BINL definition of utterance.[1] Words borrowed from the non-test language were also excluded from the count. Utterances were then scored for length in words and a "complexity index" was calculated. Results of the BINL scoring are reported in Berdan & Garcia (1982).

In addition to being coded for length and language (English or Spanish), each utterance was subsequently coded for several discourse related characteristics that seemed likely to relate to length of utterance. These variables represent a variety of attributes of language or communication. They range from characteristics that will generally be considered discourse characteristics, including the status of the speaker, and the function of turns in the conduct of the discourse. They also include more directly syntactic measures, such as the number of clauses in an utterance, and a measure that is syntactic, but dependent on the syntax of the previous turn in the discourse, ellipsis. This set of variables and the values by which utterances were categorized are listed below. They will be referred to here collectively, and somewhat loosely, as discourse variables. They included the following:

---

[1]The BINL is scored in terms of mean number of words per language sample and the complexity of the language used. Eliminated from the word count are things such as repetitions, corrections, fillers, and words substituted from another language. Borrowed words, i.e., vocabulary incorporated from another language, are counted, as are proper names. Contractions are counted as two words.

1. Ellipsis. This variable is coded for whether the child's utterance is an elliptic form of the previous relevant utterance:

   a. Completely elliptical

   b. Partially elliptical

   c. Not elliptical (but ellipsis possible)

   d. Ellipsis not appropriate (by the rules of discourse)

2. Discourse function of the turn.

3. Discourse function of the previous turn. The two variables coding discourse function used the same set of values, based on informational content and the effect of the utterance in the interaction. It is comparable to other function classification systems (e.g., Sinclair and Coulthard, 1975:40-44; Dore, 1979:354-355; Peters, Ostman, Larsen, & O'Connor, 1982).

   a. Agreement or disagreement response. An utterance which contradicted or agreed with what was said in the last relevant utterance.

   b. Request for specific information. A question which required a point of fact or conjecture from the next speaker.

   c. Elaboration. An utterance which advanced the narrative or added new information to a previous point in the discourse.

   d. Information response. An utterance which provided the specific factual or conjectural information requested by the previous speaker.

   e. Solicitation. An utterance which was procedural in nature, selecting the next speaker or otherwise indicating that the next turn be taken.

   f. Request for clarification. An utterance which required the speaker of the previous relevant turn to clarify or repeat what was said in that turn.

   g. Bid or attention getter. An utterance which was either a bid for the floor, joke, or other means of getting attention in the interaction.

h. Evaluative remark. An utterance which gave positive or negative reinforcement to the last relevant utterance.

i. Clarification. An utterance clarifying what the speaker or another speaker had said previously.

j. Request for elaboration. An utterance in which the speaker asked or otherwise indicated that another speaker give new information about the topic under discussion.

4. Syntax of the utterance.

5. Syntax of the previous turn. Syntax was grossly indicated as the number of clauses.

   a. Less than one full clause

   b. One full clause.

   c. Two clauses.

   d. Three clauses or more.

6. Speaker of the previous turn. This variable classified the speaker who had the last relevant utterance in the interaction to which the child's utterance was a response.

   a. Sibling of the child.

   b. Peer of the child.

   c. Usual fieldworker.

   d. Companion fieldworker.

   e. Adult relative.

   f. No relevant previous turn.

The one-way tabulations on each of these variables show that the distribution of utterances across is highly skewed (Table 1). For example, for Function of Turn, there were 584 examples of Elaboration, but only six examples of Request for Elaboration across all sixteen children. Cross tabulation of all variables produced numerous empty cells. Because of this, the number of categories was reduced for each variable by conflation of logically similar categories, as also shown in Table 1. This recoding formed the basis for the analyses done here.[2]

Excerpts of approximately ten pages were then selected from the transcript of each session in each language, which contained an average of 45 turns per focal child. These were submitted to five bilingual "judges" who were asked to assess the overall English language and Spanish proficiency of the child on a scale of one to ten.

Results of Analysis of Discourse Variables

The questions underlying this study, and the nature of the data set, require several different analytic approaches. We want in the first place to determine which of the discourse variables relate to differences in utterance length. We then want also to separate out effects on length due to discourse context, from the more general utterance length characterization of each child in both Spanish and English.

Natural language data sets tend to be plagued by a number of distributional characteristics which call in question the appropriateness of some conventional statistical procedures, such as analysis of variance. In the present case the chief problems are the numerous empty cells and the grossly unequal number of observations per cell. The variables which seem to be most meaningful in differentiating the effects of discourse context cross classify in such a way that there is extremely low probability of occurrence of utterances in some cells, while in other cells, utterances occur at high frequency.

[2] It may be noted in Table 1 that in recoding, the values Elaboration and Information Response were not conflated for Function of Turn but were for Function of Previous Turn. Elaboration and Information Response are themselves quite different with respect to length, but do not seem to differentially affect the subsequent turn.

Table I

Distribution of Utterances by Discourse Variables

Variable	Value	N	Mean Length	Recoded Value
Language	English	635	5.18	English
	Spanish	660	4.71	Spanish
Ellipsis	Partial	26	4.12	Ellipsis
	Full	473	3.48	
	No ellipsis	75	5.01	No ellipsis
	Not appropriate	721	5.91	
Function of Turn	Request Specific Information	130	4.85	Request Information
	Request Elaboration	6	4.67	
	Request Clarification	27	2.81	
	Elaboration	584	6.39	Elaboration
	Information Response	410	3.38	Information Response
	Clarification	54	3.42	Attention to interaction
	Agree--Disagree	43	4.74	
	Bid-Attention Getter	24	3.54	
	Evaluative remark	17	5.18	
Function of Previous Turn	Request Specific Information	442	3.57	Request Information
	Request Clarification	52	3.73	
	Request Elaboration	326	5.89	Prompt
	Solicitation	145	7.48	
	Elaboration	90	5.17	Elaboration/ Information
	Information Response	94	4.26	
	Clarification	24	4.17	Attention to Interaction
	Agree-Disagree	52	5.21	
	Bid-Attention Getter	19	6.47	
	Evaluative Remark	51	5.20	
Number of Clauses of Turn	Less than 1	474	2.31	Less than 1
	1 full clause	644	5.16	1 full clause
	2 clauses	143	10.34	2 or more clauses
	3 or more clauses	34	15.18	
Number of Clauses of previous Turn	Less than 1	434	5.30	Less than 1
	1 full clause	703	4.71	1 full clause
	2 clauses	125	4.93	2 or more clauses
	3 or more clauses	29	5.59	
Speaker of Previous Turn	Usual fieldworker	713	5.69	Fieldworker
	Companion fieldworker	212	2.84	Other adult
	Adult relative	17	4.24	
	Sibling/Peer	353	4.73	Peer

We use the following strategy to attempt to circumvent these problems. First each of the six linguistic variables is treated separately, one at a time by analysis of variance. Lack of significant effect for any of the variables suggests that it may be ignored in consideration of the overall relationship of discourse context to length. Because the variables are analyzed separately, however, there ?s the possiblity that any one of them may simply be a recoding of some other variable, e.g., an effect related to Request for Information-as-a-Function of the Previous Turn may be indistinguishable from an effect related to Information Response as a function of the turn itself. Thus these analyses of variance allow the discard of irrelevant variables, but do not identify redundant variables. To do that it is necessary to look at the remaining discourse variables simultaneously. We use two separate procedures to do that: multiple regression, and a multinomial maximum likelihood estimate. For each of these (but not for the analyses of variance) data are aggregated across children. The multiple regression, of course, reduces essentially to analysis of variance, but is the more conventional form for the subsequent calculation of residuals of length in each discourse context, a calculation which will subsequently be used as a weighting device.

The maximum likelihood estimates provide parallel information, but with somewhat different assumptions, as detailed below. Rather than just predicted mean lengths for each discourse context, the multinomial procedure yields expected frequency distributions of all lengths for each discourse context. This difference is a shift from estimating length to estimating frequencies of occurence of each length. This overall approach of successive analyses is similar in many respects to that used earlier for analysis of primary linguistic variables in Berdan (1975) and in Garcia (1981).

<u>Analyses of Variance</u>. For the analyses of variance, children were grouped according to school grade level: Preschoolers, First Graders and Third Graders. Utterances for each child were grouped separately by

Table 2

Mean Length of Utterance in English and Spanish by Discourse Variable

	English	Spanish
1. Ellipsis		
Ellipsis	3.10	3.85
No Ellipsis	4.36	5.28
2. Function of Turn		
Request Information	3.69	4.07
Elaboration	5.36	5.86
Information Response	3.26	3.66
Attention to Interaction	3.66	4.03
3. Number of Clauses		
Less than 1	2.57	2.32
Equal to 1	5.31	4.57
More than 1	3.51	2.92
4. Function of Previous Turn		
Request Information	3.34	3.74
Prompt	4.88	5.93
Elaboration Information	4.08	4.31
Attention to Interaction	4.13	4.86
5. Speaker of the Previous Turn		
Fieldworker	4.09	5.07
Other adult	3.17	4.39
Peer	4.58	3.89
6. Number of Clauses of Previous Turn		
Less than 1	5.08	4.50
Equal to 1	4.69	4.12
More than 1	4.70	4.24

language, English or Spanish. Then each of the remaining six discourse variables was treated in turn in a grade x language x discourse variable design. For each analysis, utterances were reclassified by the levels of the discourse variable in the analysis. Mean MLUs across grade levels are shown for each discourse level of each of the variables in Table 2. The significant effects ($p < .05$) for each of the six analyses are summarized in Table 3.

Table 3

Summary of Significant Effects ($p \leq .05$) for Six Analyses of Variance (Grade x Language x Discourse Variable)

VARIABLES (Levels)	MEAN SQUARE	df	f	p
ANALYSIS 1: ELLIPSIS				
Ellipsis (2)	27.363	1,12	36.40	.0001
ANALYSIS 2: FUNCTION OF TURN				
Grade (3)	26.801	2,12	5.29	.0225
Function (4)	23.283	3,36	20.25	.0000
ANALYSIS 3: NUMBER OF CLAUSES				
Clauses (3)	42.384	2,24	16.20	.0000
ANALYSIS 4: FUNCTION OF PREVIOUS TURN				
Language (2) x Grade (3)	41.493	2,12	4.05	.0452
Function (4)	14.718	3,36	9.28	.0001
ANALYSIS 5: SPEAKER OF PREVIOUS TURN				
Speaker (3)	3.010	2,24	4.30	.0254
Language (2) x Speaker (3)	6.788	2,24	5.66	.0097
ANALYSIS 6: NUMBER OF CLAUSES, PREVIOUS TURN				
Grade (3) x Language (2) x Clauses (3)	5.695	4,24	4.01	.0125

Five of the six discourse variables show significant effect on the length of utterance:

Ellipsis
Function of turn
Number of clauses
Function of the previous turn
Speaker of the previous turn

The sixth variable, Number of Clauses of the Previous Turn, showed no significant effect. As shown in Table 3, there is an effect for grade level, but only in the analysis by Function of Turn. In none of the analyses is there a significant main effect for language. There are several interactions, but for none of them are all of the relevant main effects significant. Inspection of the cell means suggests that there is some tendency for shorter utterances by preschool children, particularly in English. In general, the preschool children did not use multiple clause utterances in English. However, given that the six analyses are simply reclassifications of the same data set, grade and language effects, which are not consistent across analyses, are highly suspect.

These analyses lend strong support to our general contention that discourse context is an important intervening variable in the interpretation of the relationship of MLU to language development. Analyzing the discourse variables separately, however, entails the possibility that one or more of the observed effects is nothing more than a re-labelling of some other logically prior effect. Considering the discourse variables simultaneously in this data set, however, introduces the distributional problems referred to above. For this reason we turn first to multiple regression, and then to a maximum likelihood procedure.

<u>Multiple Regression Analysis.</u> In the multiple regression analysis we wish to look at the effects of all variables simultaneously. In order to maintain the size of each cell, we aggregated the observations

across speakers. The two syntactic variables, Number of Clauses and Number of Clauses in the Previous Turn, were not used in this analysis. The analysis of variance showed no effect on length for Number of Clauses of the Previous Turn. Number of Clauses for the child's own turn did show a significant effect (Analysis 3). However, Number of Clauses is not, strictly speaking, an independent variable, but an alternate to number of words as a representation of length of utterance. It correlates highly with length measured in words ($r = .75$), and virtually precludes showing any other significant effects in multiple regression.

The four remaining discourse variables: Ellipsis, Function of Turn, Function of Previous Turn, and Speaker of Previous Turn; with Language as a fifth, are the independent variables used for the multiple regression. For regression, each of these variables is construed as a nominal scale. The values of these scales were dummy coded according to the procedures outlined in Cohen & Cohen (1975:175 ff.) In this coding each value becomes a dichotomous variable; each utterance is scored according to "presence" or "absence" of each value of each variable. In this scoring, one value for each variable becomes in one sense redundant, and in another sense becomes a reference point for comparison to the other values. This status was given arbitrarily to the "Attention to Interaction" values for the Function variables, and to the "Other Adult" value of the Speaker of the Previous Turn. Ellipsis and Language were already dichotomous variables.

The results of the multiple regression analyses are shown first for each discourse variable separately, analogous to the analysis of variance treatment. For these, R^2 represents the proportion of variance in length attributable to each variable alone, when the other variables are not considered. These are shown in Table 4. Values of R^2 range from .007 for Language, to .150 for Function of Turn. All are significant beyond $\alpha = .01$, by conventional F tests. However, the F ratios should not be interpreted literally, since the degrees of freedom for their denominators represent all utterances aggregated across all children.

Table 4

Multiple Regressions for Each Discourse Variable Treated Separately

INDEPENDENT VARIABLE	BETA	MULTIPLE R	R^2
ELLIPSIS		.339	.115
Ellipsis	-.339		
FUNCTION OF TURN		.387	.150
Request Information	.054		
Elaboration	.309		
Information Response	-.119		
SPEAKER OF PREVIOUS TURN		.303	.092
Fieldworker	.411		
Peer	.256		
FUNCTION OF PREVIOUS TURN		.356	.127
Request Information	-.234		
Prompt	.165		
Elaboration/Information	-.026		
LANGUAGE		.082	.007
English	.082		

The sets of dummy variables representing each discourse variable were then entered into the regression step-wise, in the order that they are listed in Table 4. This ordering was not based strictly on logical precedence among the variables, but on the relative magnitude of effects in the analyses of variance, and on the ease with which the variables can be coded. This latter consideration confounds other orderings, but is of considerable interest if this procedure is to have practical application. In particular, it is generally easier to identify the speaker of the previous turn than it is to code the discourse function of that turn. Other orderings than the one presented here merit consideration. In Table 5 the values of Beta are given for each dummy variable, for each step that variable is included in the regression equation. The resulting value for R^2 and the change in R^2 for each step are also shown.

Table 5

Step-wise Multiple Regression

VARIABLES	VALUES OF BETA STEP 1	STEP 2	STEP 3	STEP 4	STEP 5
ELLIPSIS	-.339	-.225	-.198	-.213	-.232
FUNCTION OF TURN					
Request Info		.005	.022	.017	-.004
Elaboration		.273	.283	.236	.223
Info Response		-.005	.046	.057	.048
SPEAKER PREV TURN					
Fieldworker			.285	.269	.248
Peer			.212	.251	.244
FUNCTION PREV TURN					
Request Info				.037	.021
Prompt				.132	.127
Elab-Info				-.017	-.021
LANGUAGE					.120
R^2	.115	.177	.218	.227	.240
R^2 Change	.115	.062	.041	.009	.013
F(R^2 Change)	166.684	32.481	33.261	4.848	23.036
df	1,1284	3,1284	2,1284	3,1284	1,1284

At each step the change in R^2 is significant, $p < .01$. The step-wise regression shows that much of the influence of Function of Turn can also be accounted for by Ellipsis. Once the Speaker of the Previous Turn is identified, less than 1% additional variance is accounted for by also including the Function of the Previous Turn. This small increment is in spite of the relatively large R^2 for Function of the Previous Turn ($R^2 = .127$) when that variable is considered in isolation. The contribution of language is extremely small in either case, but is slightly increased when the effects of the other discourse variables are also considered. As could be expected, the effects of the discourse variables are neither fully independent nor completely redundant. Total R^2 is 0.24, suggesting that about one-fourth of the total variance in utterance length can be accounted for by differentiated discourse contexts.

Maximum likelihood estimates. An alternative means of considering all of the discourse variables simultaneously is provided by maximum likelihood estimation. For this approach, utterances are tallied, and the data of interest are the frequencies with which utterances of each length occur in each discourse context. The distribution of lengths will here be regarded as a multinomial function. Estimation of such multinomials by maximum likelihood is discussed by Edwards (1972).

Sociolinguists who have used quantified approaches to the study of natural language variability have long been frustrated by the distributional characteristics of most language variables, with many contexts of great interest occurring naturally at very low frequencies (cf. Labov, 1966). The use of maximum likelihood techniques for estimating effects of linguistic environments was introduced by Cedergren (1973) and Cedergren & Sankoff (1974). The several models which these researchers estimated by maximum likelihood have more recently been replaced by a model based on logistic transformation (Cox, 1970; Lindsey, 1975) of proportions (Rousseau & Sankoff, 1976). This model has been extended to the description of polychotomous variables by Jones (1975). The introduction of this latter model to the treatment of language variation follows the development of the necessary computer programming by P. Rousseau at the University of Montreal.[3]

Terminology in the related sociolinguistic literature diverges from that of other statistical treatments. "Factor group" and "factor" are used analogously to the analysis of variance terms "variable" and "level," respectively. Within a factor group, factors give a mutually exclusive and exhaustive characterization of all observations. Any given observation is defined by one factor from each factor group and, in this case, by a category corresponding to length.

To apply this procedure here, we model length of utterance as a multinomial function, classifying utterances by length in words. In order to reduce the number of parameters estimated in the multinomial and to balance the data set, length was recoded into the following ten categories:

	Length Category									
	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
No. of words	1	2	3	4	5	6	7	8-9	10-11	12-31

This results in a compression of the upper end of the length scale, but affects a relatively small proportion of the data set.

In a sense, treatment as a multinomial degrades the information in the data set, since length is treated as a nominal rather than interval scale. However, the treatment of length as an ordinal scale is itself troublesome, since longer utterances do not result simply from

---

[3]We are grateful to Dr. Pascale Rousseau for providing us with a copy of this program and related documentation.

incrementing shorter utterances, but from a complex change in the semantic and syntactic integration of information. Whatever limitations there may be in the linguistic or cognitive interpretation of a multinomial model of length, other models of this secondary language measure seem at least equally opaque, and do not share many of the useful properties of this approach.

Thus, data are characterized as a frequency distribution. The frequency with which observations fall into each of the cells is defined by one factor from each factor group, and by the relevant length category. The cells form a matrix of $n + 1$ dimensions, where there are n factor groups. It is convenient, however, to represent the observations as a two dimensional matrix, with the rows defined by the actually occurring combinations of factors, and the columns defined as the length categories. The actual number of rows in these data is considerably less in each of the analyses given here than the possible number of rows from the cross classification of all factors in all factor groups. A row in the matrix contains all observations for an occurring combination of factors. A column contains all observations of a particular length. The first few rows of the frequency distribution for Analysis 1 below are given in Table 6.

Table 6

Sample Data Display. Number of Utterances by Factor and by Length Category

	Factor Groups 1	2	3	4	5		Length Categories 1	2	3	4	5	6	7	8	9	10
	1	1	1	1	1	Number of Obervations	0	0	1	0	0	0	0	0	0	0
Factors	1	1	1	2	2		0	0	3	2	0	0	0	1	0	0
	1	1	1	3	1		62	32	14	9	6	1	2	0	0	0
	1	1	1	3	2		0	1	5	10	5	14	9	6	1	0

The object of the maximum likelihood procedures is to derive a multinomial function for each factor in each factor group, and to calculate the likelihood that these functions adequately characterize the data set. Whether or not the use of additional factors significantly improves the characterization of the data set may be tested by comparing likelihoods across solutions. Also, from the functions defining each factor it is possible to estimate the probability of occurrence of utterances of each length category in each discourse context.

From the frequency matrix, a set of coefficients is calculated for each factor. Additionally, in order to force a unique solution, one other factor is calculated which applies in common to all rows. This is referred to in the sociolinguistic studies as the "input factor"; elsewhere as "overall effect," or a "constant." These coefficients are calculated so that the probability for any cell is the sum of the logistic transforms of the n factors defining its row, plus the logistic transform of the overall effect. The numerous desirable properties of the logistic transformation for dealing with data in the form of proportions are discussed in Rousseau and Sankoff (1976), and in Hofacker (1982).

The maximum likelihood procedure uses a nonlinear iterative technique in which all coefficients are estimated simultaneously. The procedure is iterated until it converges on a maximum value for the log likelihood (Jones, 1975), or until it becomes apparent that convergence is not possible. For the analyses below which failed to converge under this algorithm, we report the probability coefficients from the iteration with the highest log likelihood.

Maximum likelihood could be estimated using all the discourse variables (factor groups), or any subset of them. From among all of the possibilities we have sampled two combinations of five factor groups, two combinations of four factor groups, and finally, we have split the data set by language, and run separate estimates for each language set using three factor groups each. Each choice of factor groups is in effect a hypothesis of the underlying dimensions of the data set. The factor groups used to define the data set in each run, and the resulting log likelihoods are shown in Table 7.

Table 7

Factor Groups Used for Each Maximum Likelihood Estimate

FACTOR GROUPS	LOG LIKELIHOOD
ANALYSIS 1:	-2263.15
Language	
Ellipsis	
Function of Turn	
Function of Previous Turn	
Number of Clauses	
ANALYSIS 2:	-2643.56
Language	
Ellipsis	
Function of Turn	
Function of Previous Turn	
Speaker of Previous Turn	
ANALYSIS 3:	-2682.03
Language	
Ellipsis	
Function of Turn	
Function of Previous Turn	
ANALYSIS 4:	-2667.16
Language	
Ellipsis	
Function of Turn	
Speaker of Previous Turn	
ANALYSIS 5a: (English data only)	-1303.92
Ellipsis	
Function of Turn	
Function of Previous Turn	
ANALYSIS 5b: (Spanish data only)	-1336.88
Ellipsis	
Function of Turn	
Function of Previous Turn	

In spite of the difficulty of interpreting Number of Clauses as an independent variable we retained it in Analysis 1, but not in the subsequent treatments. Analyses 3 and 4 alternate between Function of the Previous Turn and Speaker of the Previous Turn. In each of these estimates, Language (Spanish and English), was also maintained as a factor group, in spite of the failure to show a language effect in the analyses of variance. For our ultimate purposes of modeling language development the appropriateness of aggregating across languages is open to serious question, and we prefer to demonstrate similarity (or difference) across languages, rather than assume it. For Analyses 5a and 5b, the data set was split by language, and separate estimates were run for each set. Otherwise these are identical in factor groups to Analysis 3. For each of the analyses, utterances are aggregated across speakers.

<u>Comparing Alternative Analyses.</u> Maximum likelihood solutions are generally evaluated in two different ways. One approach, which might be termed <u>absolute</u>, compares how well probability or frequency distributions estimated using the maximum likelihood solutions fit the observed frequencies. Using cumulative frequency displays, Figures 1-3 show observed distributions and those predicted by the maximum likelihood estimates for the first three rows of data given in Table 6. Conventional measures of goodness of fit, particularly χ^2, are difficult to interpret for these data, where there are many cells with very low expected frequencies, and no observations at all in many of the possible cross-classifications of factors.

An alternative to such absolute tests of goodness of fit are <u>relative</u> tests, comparing alternative solutions. Solutions are preferred which both maximize the log likelihood, and minimize the number of parameters estimtaed from the data set. This is tested by computing twice the difference in log likelihoods for any two solutions, and comparing that figure with the χ^2 distribution, with degrees of freedom equal to the change in degrees of freedom between the two solutions.

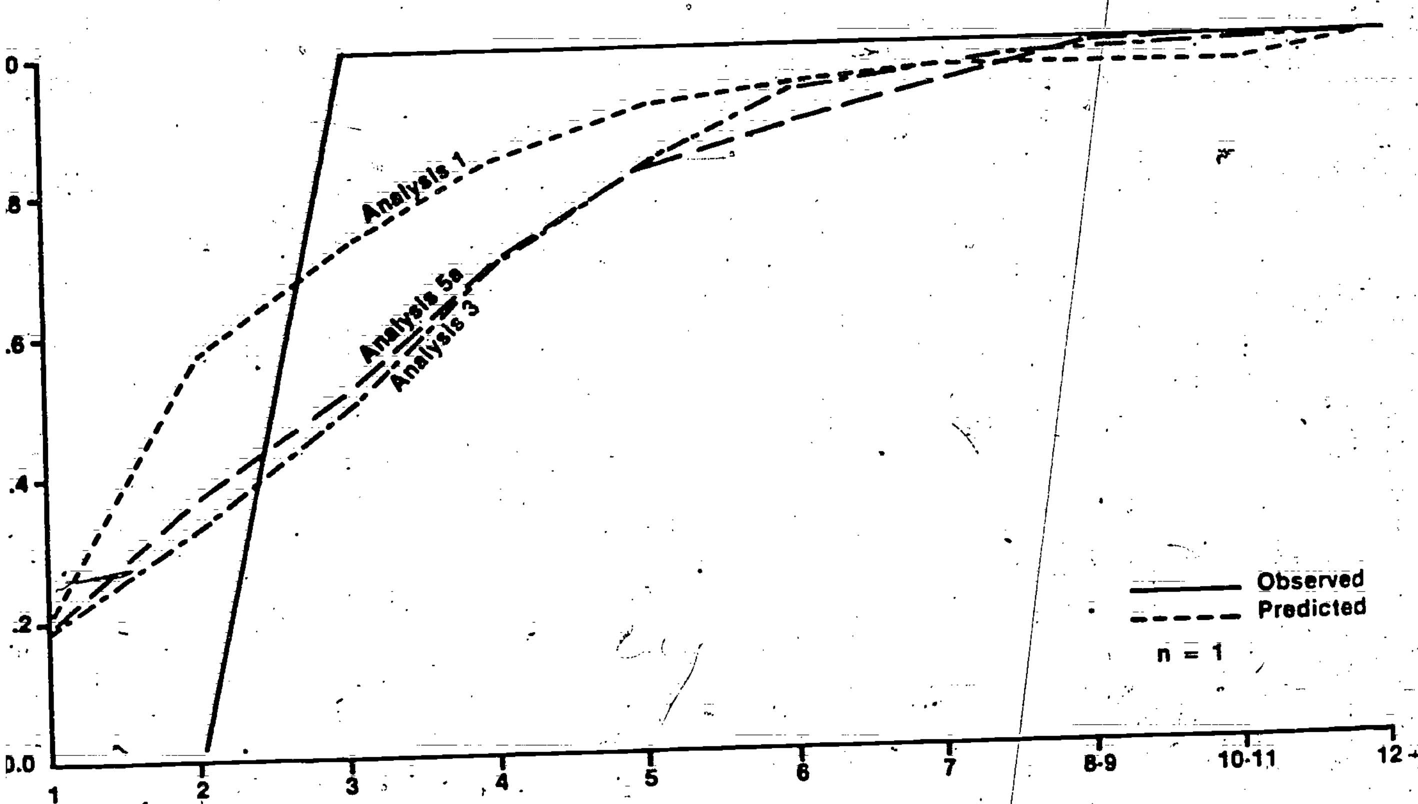

Figure 1. Cumulative frequency distribution (%) factors:
Analysis 1: English, ellipsis, request information, partial clause.
Analysis 3: English, ellipsis, request information, request information.
Analysis 5a: Ellipsis, request information, request information.

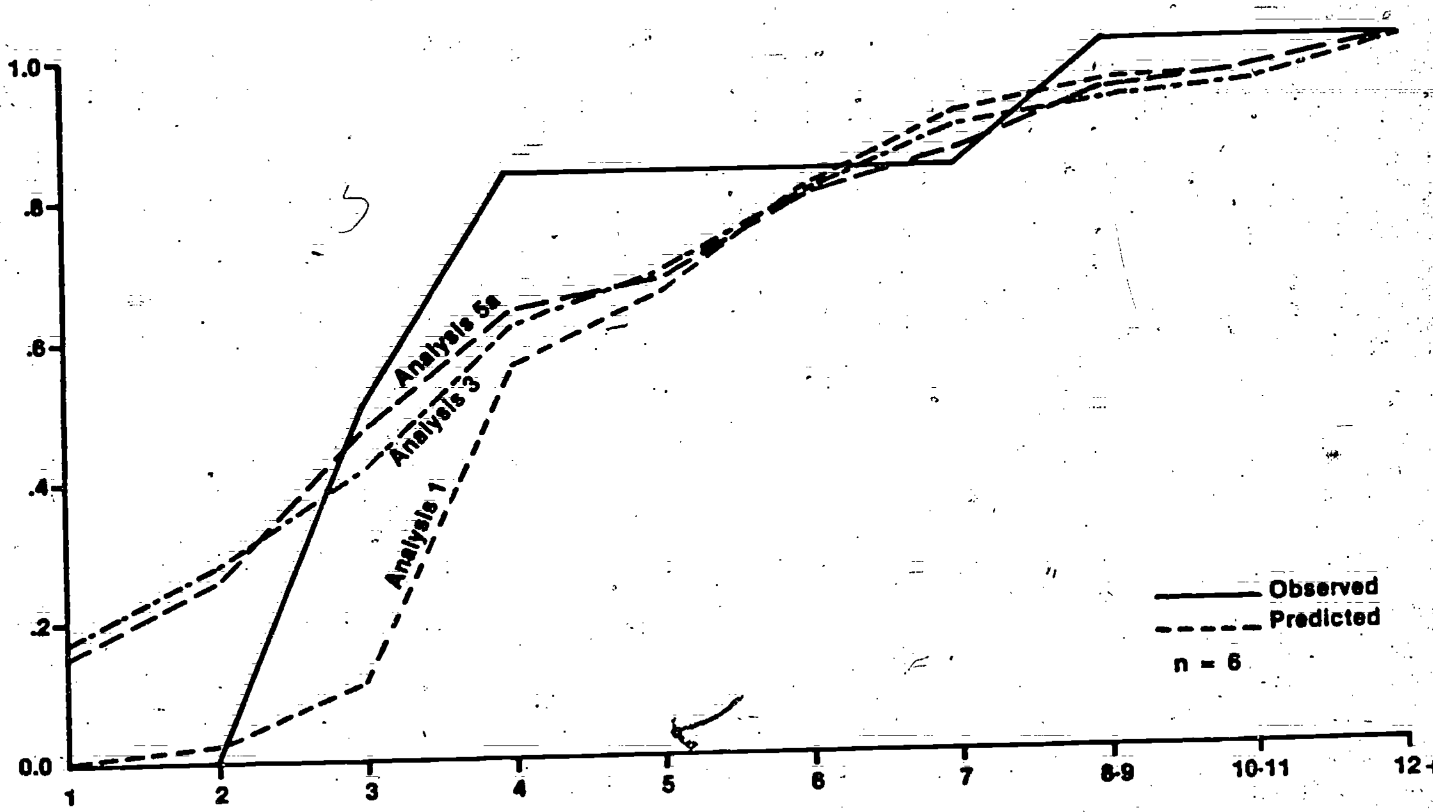

Figure 2. Cumulative frequency distribution (%) factors:
Analysis 1: English, ellipsis, request information, elaboration, one clause.
Analysis 3: English, ellipsis, request information, elaboration.
Analysis 5a: Ellipsis, request information, elaboration.

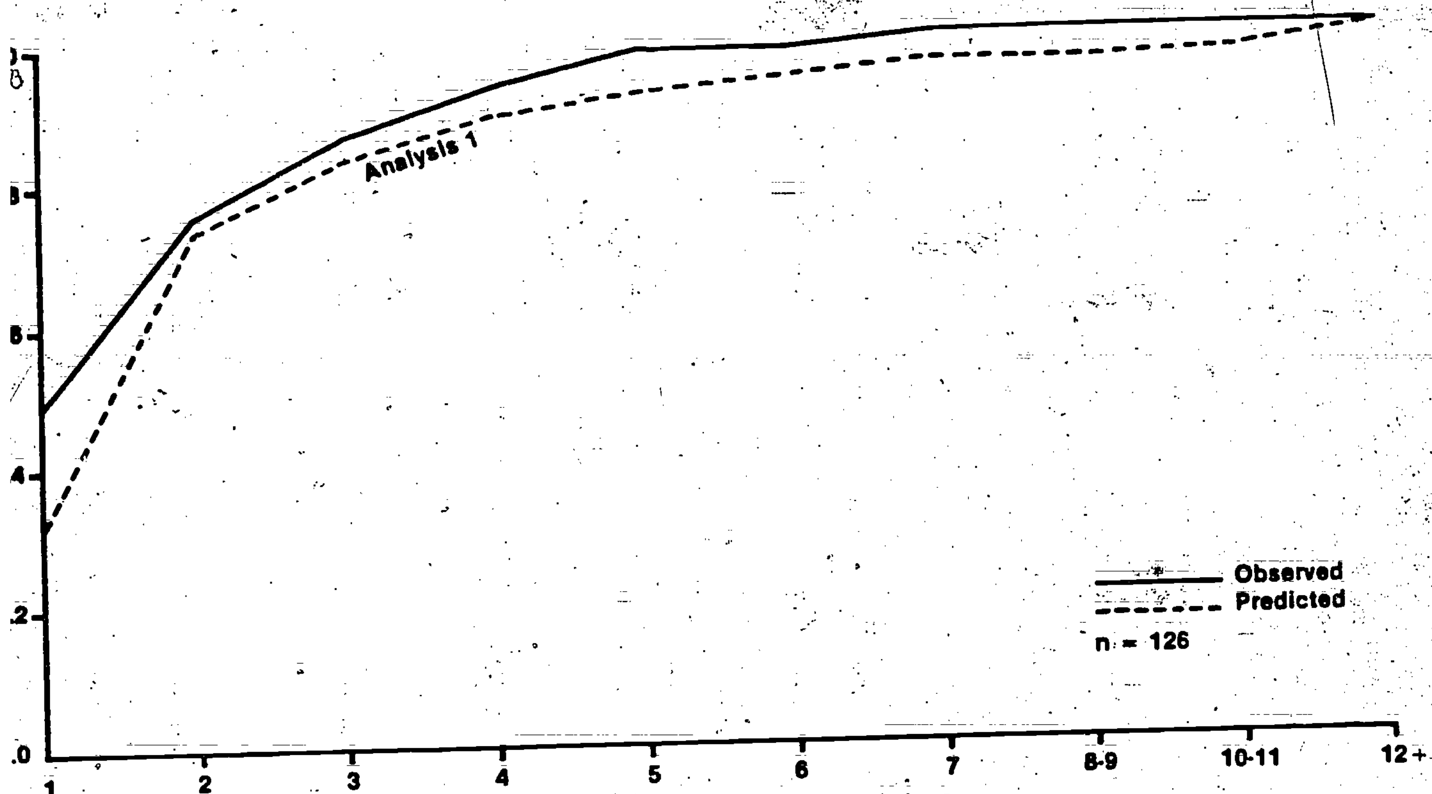

Figure 3.1. Cummulative frequency distribution (%) factors: Analysis 1: English, ellipsis, request information, information response, partial clause.

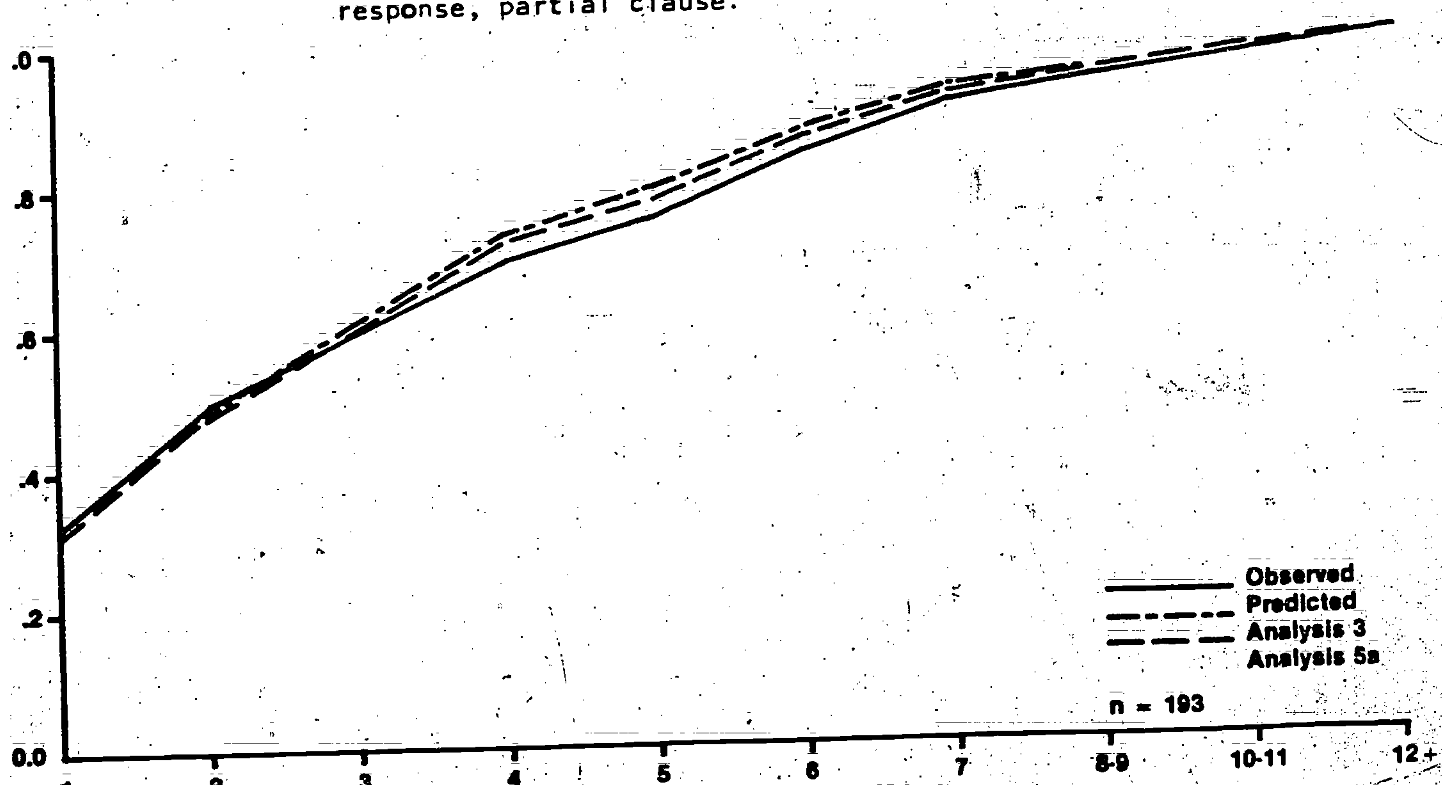

Figure 3.2. Cumulative frequency distribution (%) factors: Analysis 3: English, ellipsis, request information, information response. Analysis 5a: Ellipsis, request information, information response.

By this criterion, first choice among the solutions is Analysis 1, incorporating Number of Clauses as a factor group. Its log likelihood is significantly greater than that of any other solutions. Analysis 2 is significantly better than Analysis 3 (χ^2 (2,9) = 79.94), but not better than Analysis 4 (χ^2 (3,9) = 23.60). Splitting the data set by languages in Analyses 5a and 5b nearly doubles the number of parameters estimated, but does not significantly improve the log likelihood (χ^2 (6,9) = 41.23) over Analysis 3.

With the multinomial approach used here it is possible to examine the effect each factor across the length categories. The coefficients estimate for each factor in each analysis are given in Table 8. They are also shown graphically for Analyses 1, 3, 5a, and 5b, in Figures 4-7, respectively. Coefficients of 0.1 (the reciprocal of the number of categories in the multinomial) have no effect; larger coefficients lead to estimates of higher probabilities, smaller coefficients to lower probabilities.

The three different treatments of the data set shown graphically yield generally different patterns of coefficients for the factors, and some patterns that seem quite consistent across treatments. In Analysis 1, Overall Effect shows three separate highs, which seem to result from classifying the utterances by number of clauses. In Analysis 3, Overall effect is a generally decreasing function of length, roughly comparable to the overall frequency distribution of length. In Analysis 5a and 5b, where the data set is split by language, the general shape of Analysis 2 is maintained, but with opposite effects at Lengths 3 and 4.

<u>Language Factor.</u> The overall shape of the effects for the Language factors is the same in Analyses 1 and 3, but of considerably greater magnitude in Analysis 1. In both cases, Lengths 1-3 are more probable in English than in Spanish, with the opposite effect at the high end of the length scale. Although language seems to show a fairly small effect, particularly when number of clauses is not considered (Analysis 3), the separate analyses in 5a and 5b show quite different effects for some factors.

Table 8: Coefficients for Factors

Analysis 1 Log Likelihood = -2263.15

Language	1	=	0.053	0.061	0.065	0.108	0.097	0.127	0.109	0.130	0.147	0.102
	2	=	0.170	0.147	0.138	0.083	0.092	0.071	0.082	0.069	0.061	0.088
Ellipsis	1	=	0.122	0.104	0.098	0.092	0.096	0.112	0.130	0.070	0.087	0.089
	2	=	0.079	0.093	0.099	0.106	0.102	0.087	0.075	0.139	0.112	0.109
Function of previous turn	1	=	0.098	0.084	0.127	0.159	0.097	0.094	0.072	0.082	0.076	0.111
	2	=	0.055	0.074	0.074	0.079	0.084	0.104	0.083	0.112	0.240	0.094
	3	=	0.145	0.122	0.071	0.074	0.098	0.120	0.108	0.114	0.059	0.088
	4	=	0.109	0.112	0.128	0.091	0.107	0.072	0.130	0.081	0.078	0.093
Function of turn	1	=	0.079	0.100	0.118	0.101	0.167	0.113	0.071	0.085	0.070	0.096
	2	=	0.063	0.069	0.097	0.099	0.084	0.117	0.125	0.121	0.116	0.108
	3	=	0.128	0.122	0.085	0.057	0.083	0.088	0.089	0.102	0.153	0.094
	4	=	0.139	0.105	0.090	0.157	0.075	0.076	0.113	0.085	0.071	0.091
Syntax	1	=	0.086	0.686	0.016	0.199	0.005	0.003	0.002	0.001	0.001	0.003
	2	=	0.004	0.193	0.024	0.676	0.030	0.025	0.017	0.107	0.009	0.005
	3	=	0.011	0.000	0.010	0.000	0.026	0.052	0.106	0.185	0.351	0.259
		=	0.083	0.006	0.152	0.005	0.141	0.122	0.139	0.088	0.033	0.221

Analysis 3 Log Likelihood = -2682.03

	1	=	0.069	0.075	0.072	0.113	0.097	0.125	0.104	0.119	0.124	0.102
	2	=	0.139	0.127	0.132	0.085	0.099	0.076	0.092	0.080	0.077	0.093
	1	=	0.189	0.145	0.104	0.091	0.085	0.095	0.111	0.055	0.065	0.059
	2	=	0.046	0.060	0.084	0.096	0.103	0.092	0.079	0.158	0.134	0.148
	1	=	0.107	0.084	0.123	0.146	0.092	0.097	0.073	0.085	0.091	0.100
	2	=	0.052	0.072	0.076	0.086	0.090	0.111	0.086	0.115	0.218	0.094
	3	=	0.159	0.129	0.170	0.073	0.094	0.110	0.103	0.109	0.056	0.096
	4	=	0.097	0.110	0.131	0.094	0.109	0.072	0.133	0.080	0.077	0.195
	1	=	0.088	0.118	0.131	0.114	0.182	0.119	0.071	0.078	0.057	0.042
	2	=	0.056	0.056	0.076	0.076	0.066	0.095	0.109	0.113	0.138	0.215
	3	=	0.131	0.121	0.083	0.052	0.077	0.077	0.078	0.096	0.151	0.134
	4	=	0.119	0.096	0.093	0.171	0.084	0.088	0.128	0.092	0.066	0.063
		=	0.173	0.151	0.154	0.140	0.122	0.086	0.074	0.048	0.020	0.032

Analysis 5a (English) Log Likelihood = -1303.92

	1	=	0.259	0.156	0.093	0.085	0.069	0.084	0.098	0.058	0.057	0.040
	2	=	0.030	0.050	0.083	0.092	0.112	0.092	0.079	0.134	0.137	0.192
	1	=	0.055	0.059	0.148	0.125	0.087	0.111	0.063	0.138	0.086	0.128
	2	=	0.048	0.060	0.058	0.068	0.122	0.132	0.100	0.097	0.208	0.108
	3	=	0.248	0.128	0.066	0.070	0.086	0.135	0.063	0.083	0.056	0.065
	4	=	0.105	0.154	0.124	0.118	0.076	0.035	0.176	0.064	0.070	0.077
	1	=	0.090	0.142	0.107	0.085	0.169	0.084	0.091	0.128	0.067	0.037
	2	=	0.051	0.055	0.109	0.058	0.039	0.114	0.067	0.158	0.109	0.242
	3	=	0.137	0.100	0.070	0.049	0.063	0.091	0.077	0.080	0.121	0.211
	4	=	0.098	0.080	0.076	0.253	0.149	0.071	0.131	0.038	0.069	0.033
		=	0.139	0.146	0.106	0.196	0.131	0.085	0.095	0.050	0.031	0.021

Analysis 5b (Spanish) Log Likelihood = -1336.88

	1	=	0.161	0.136	0.110	0.082	0.095	0.103	0.119	0.037	0.069	0.089
	2	=	0.053	0.063	0.078	0.104	0.091	0.084	0.072	0.235	0.124	0.096
	1	=	0.172	0.105	0.109	0.151	0.102	0.095	0.085	0.024	0.074	0.081
	2	=	0.044	0.073	0.087	0.102	0.064	0.097	0.070	0.176	0.203	0.083
	3	=	0.117	0.121	0.064	0.062	0.093	0.079	0.130	0.164	0.061	0.109
	4	=	0.087	0.082	0.126	0.080	0.126	0.105	0.098	0.108	0.084	0.103
	1	=	0.091	0.094	0.151	0.153	0.166	0.210	0.030	0.051	0.000	0.054
	2	=	0.017	0.015	0.016	0.023	0.027	0.020	0.052	0.026	0.752	0.050
	3	=	0.045	0.054	0.035	0.021	0.037	0.020	0.031	0.045	0.676	0.036
	4	=	0.076	0.067	0.060	0.071	0.032	0.063	0.108	0.088	0.379	0.054
		=	0.209	0.168	0.215	0.100	0.112	0.078	0.049	0.020	0.001	0.047

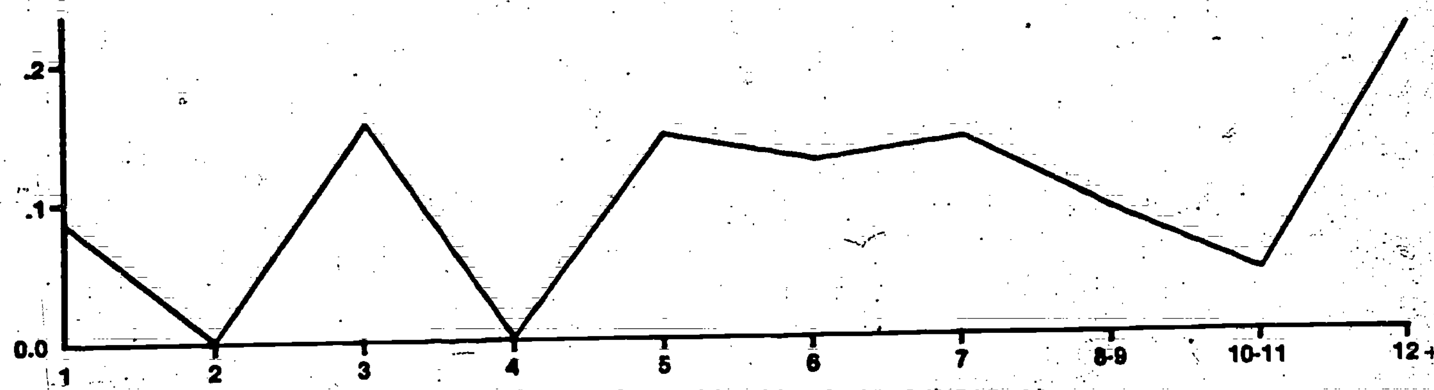

Figure 4.1. Overall effect (Analysis 1).

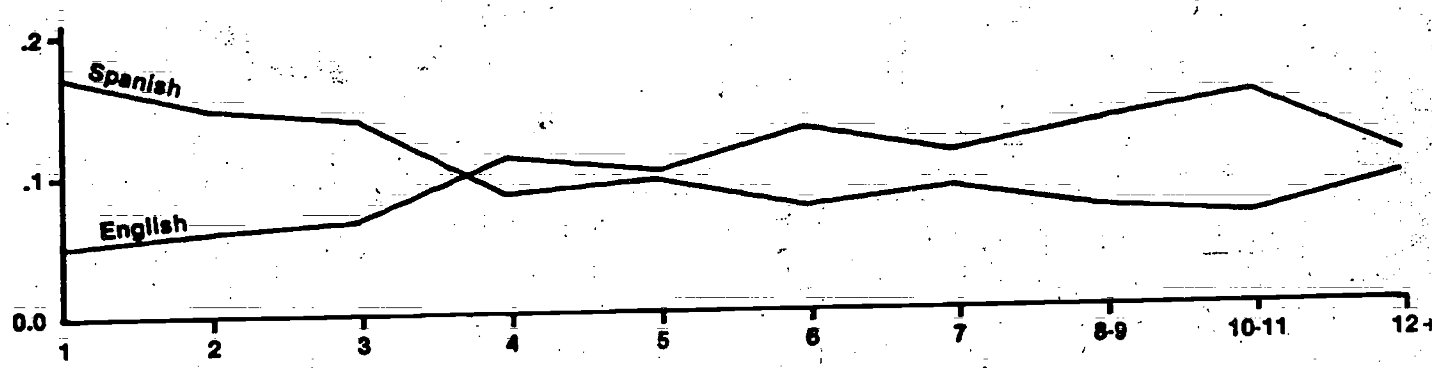

Figure 4.2. Language (Analysis 1).

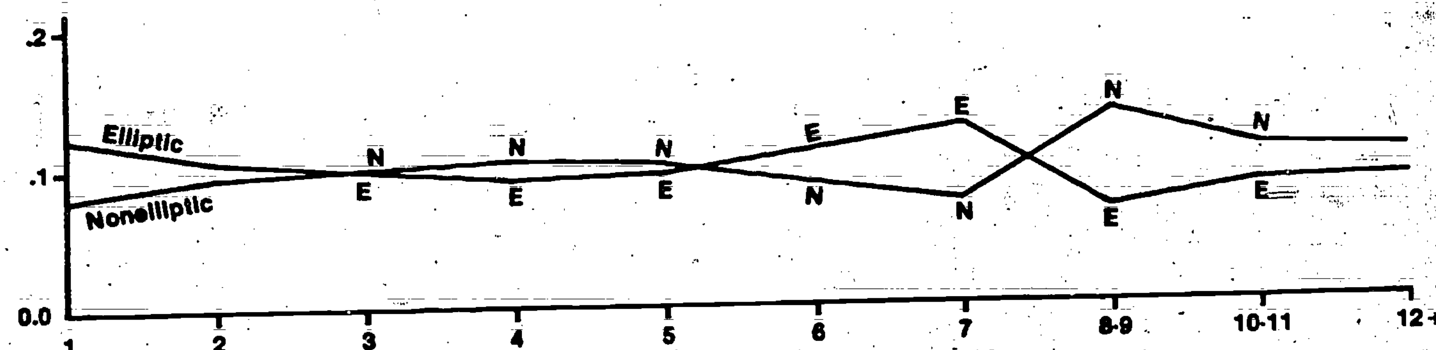

Figure 4.3. Ellipsis (Analysis 1).

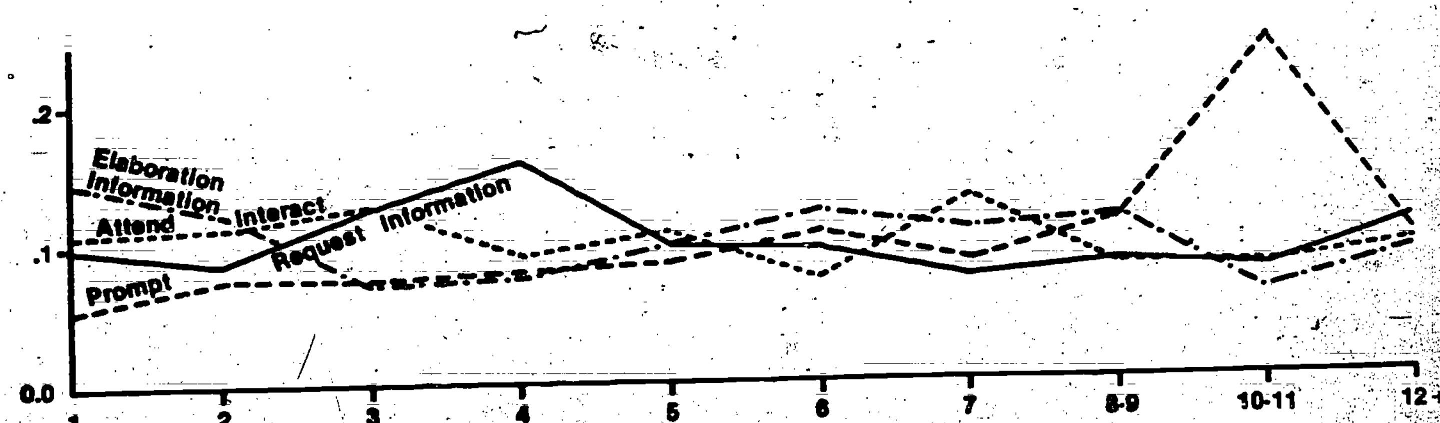

Figure 4.4. Function of previous turn (Analysis 1).

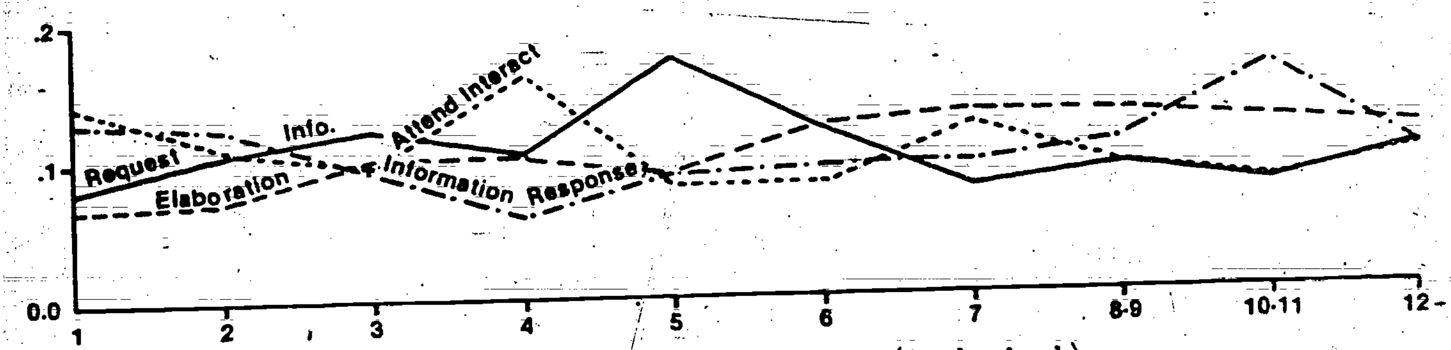

Figure 4.5. Function of turn (Analysis 1).

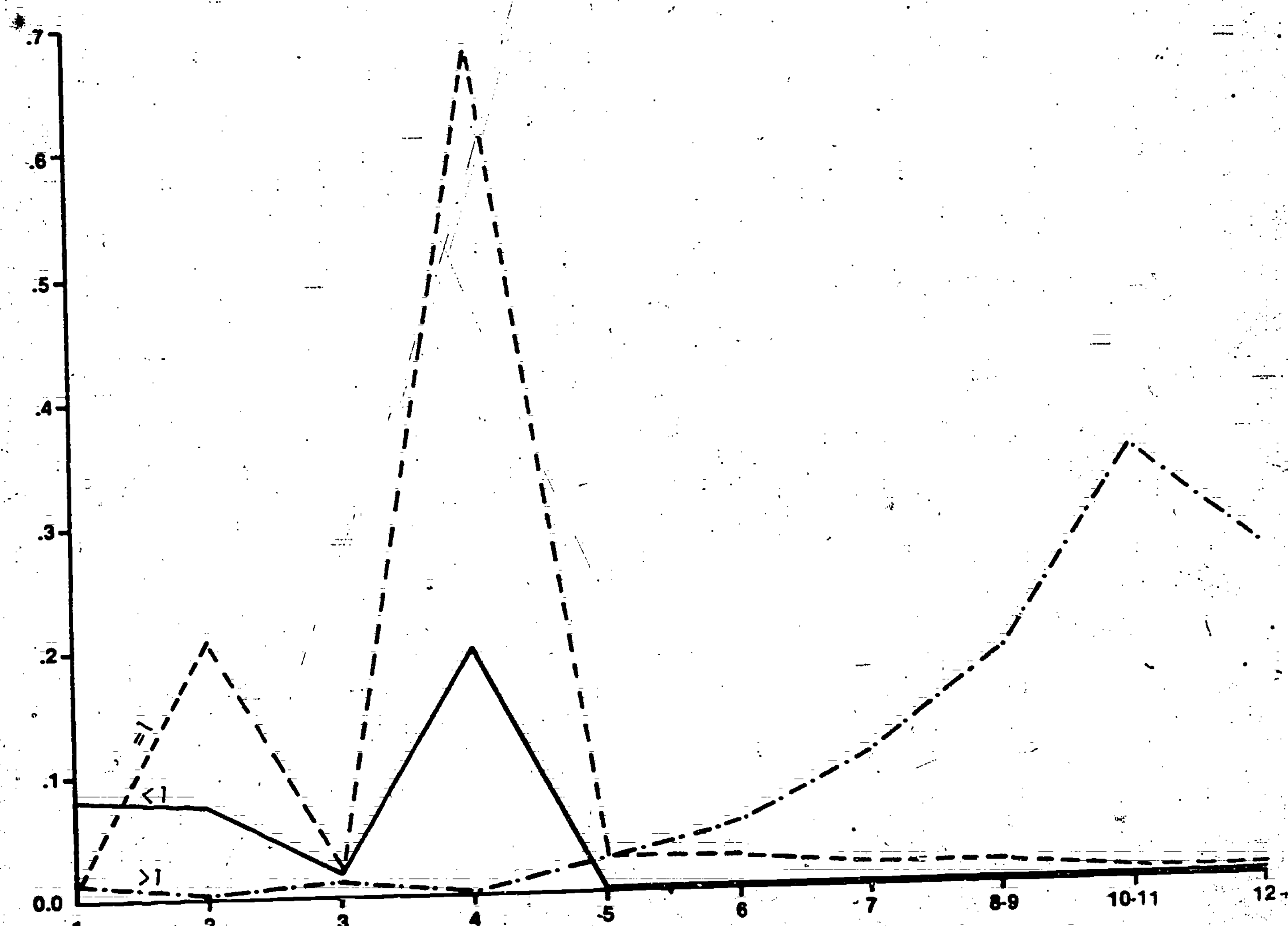

Figure 4.6. Number of clauses (Analysis 1).

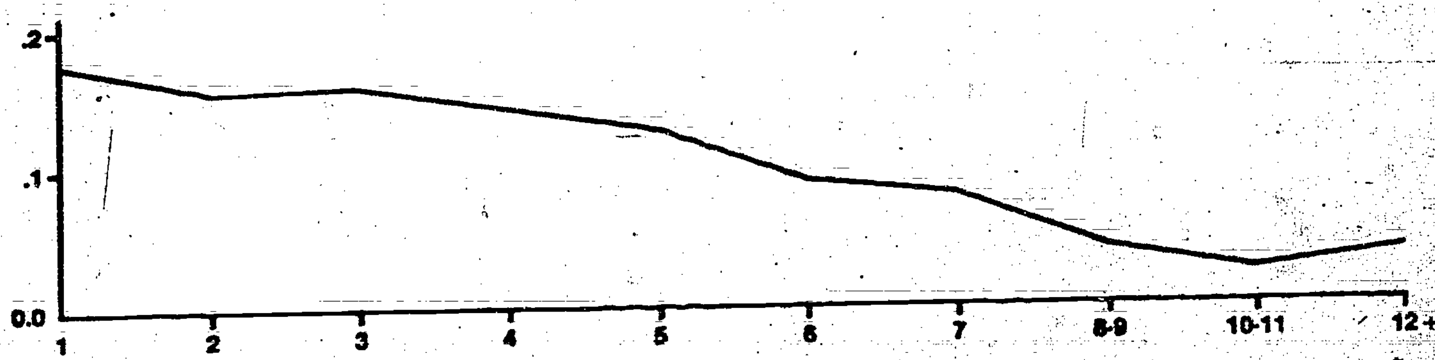

Figure 5.1. Overall effect (Analysis 3).

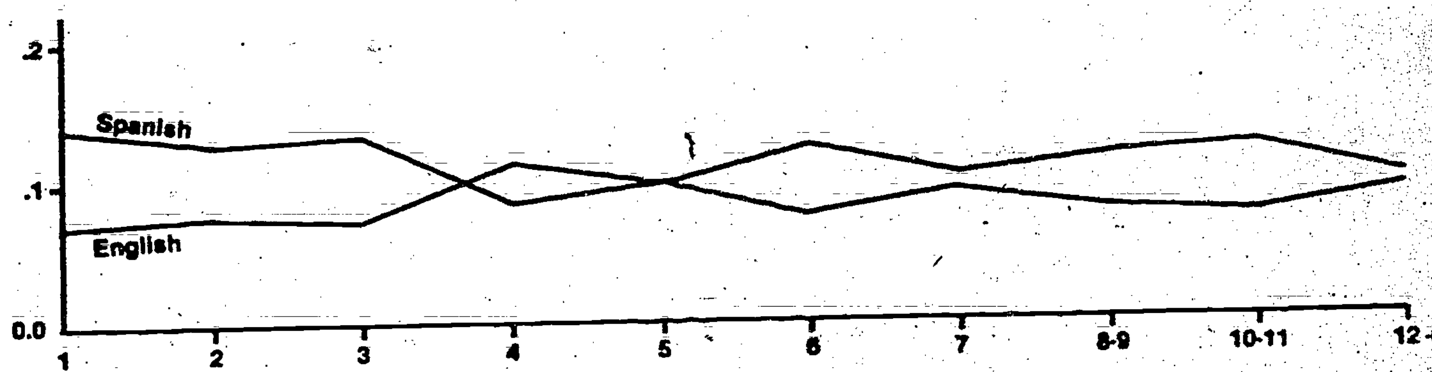

Figure 5.2. Language (Analysis 3).

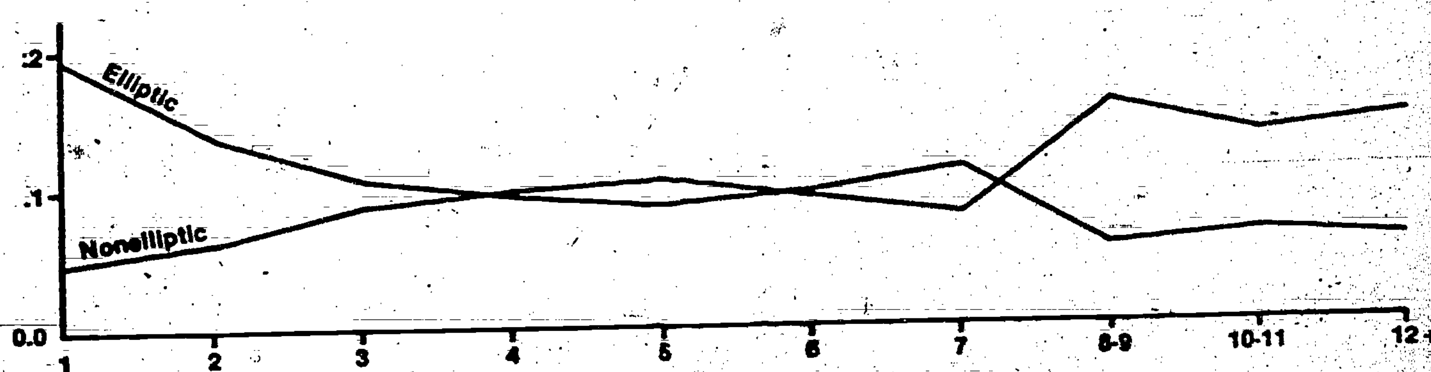

Figure 5.3. Ellipsis (Analysis 3).

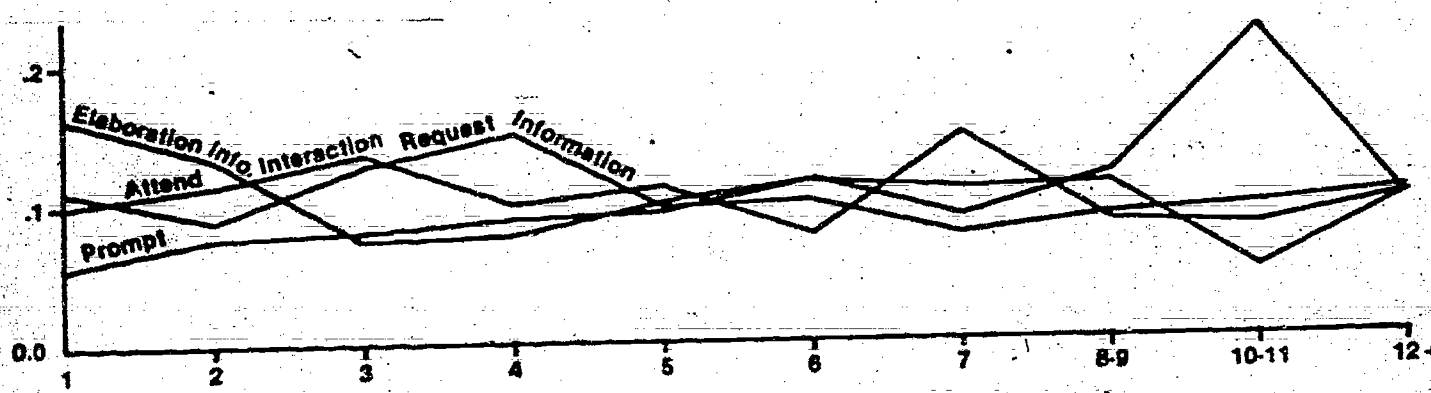

Figure 5.4. Function of previous turn (Analysis 3).

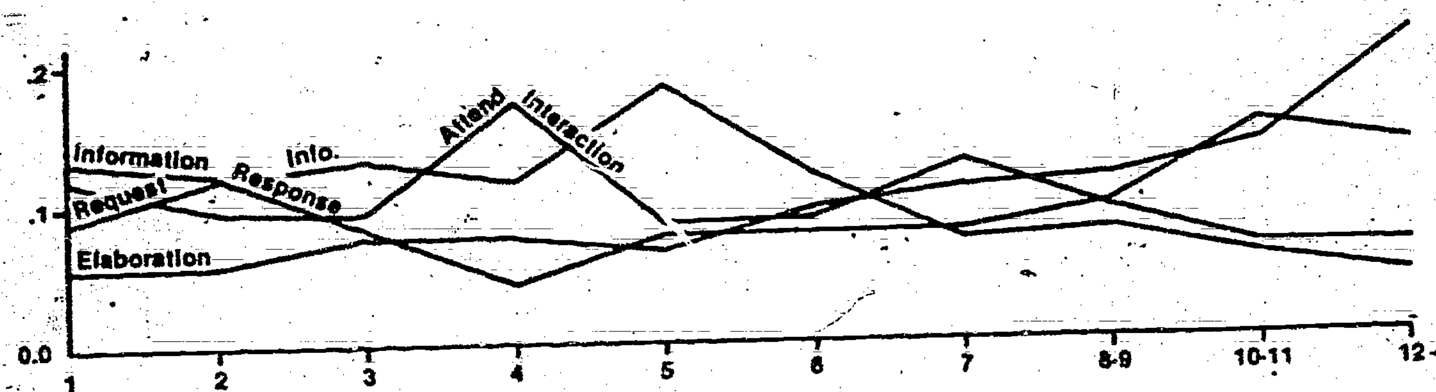

Figure 5.5. Function of turn (Analysis 3).

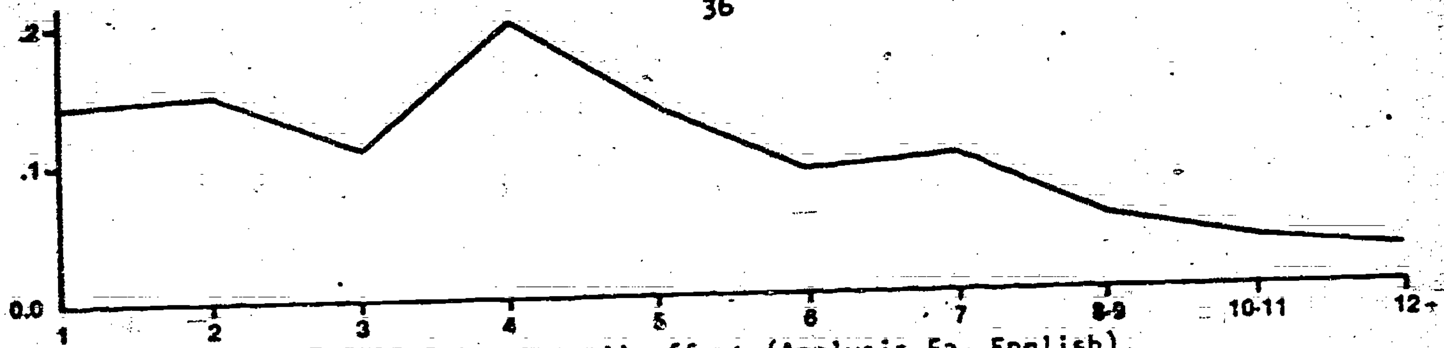

Figure 6.1. Overall effect (Analysis 5a, English).

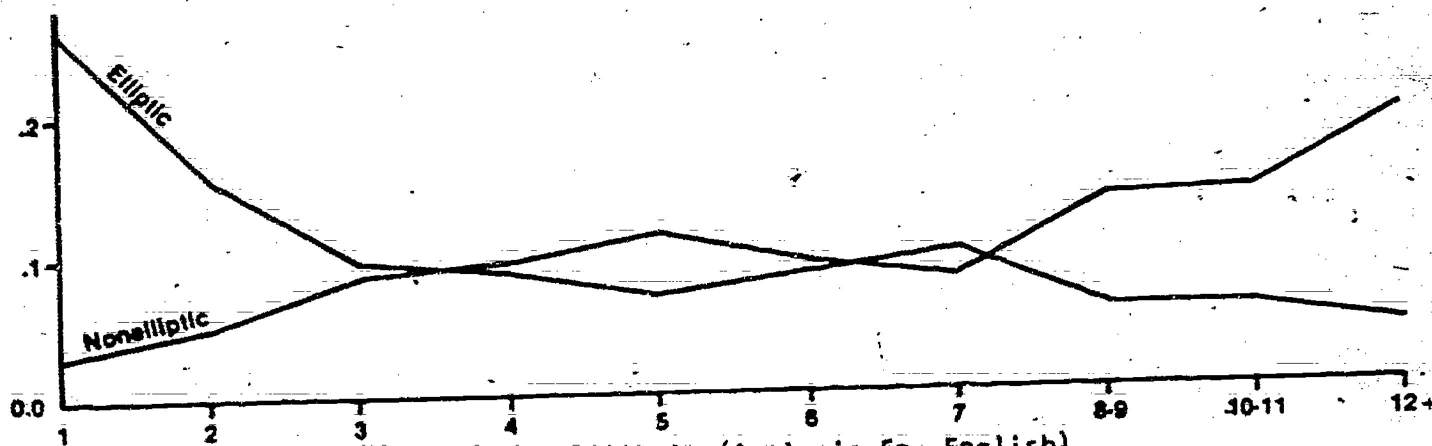

Figure 6.2. Ellipsis (Analysis 5a, English).

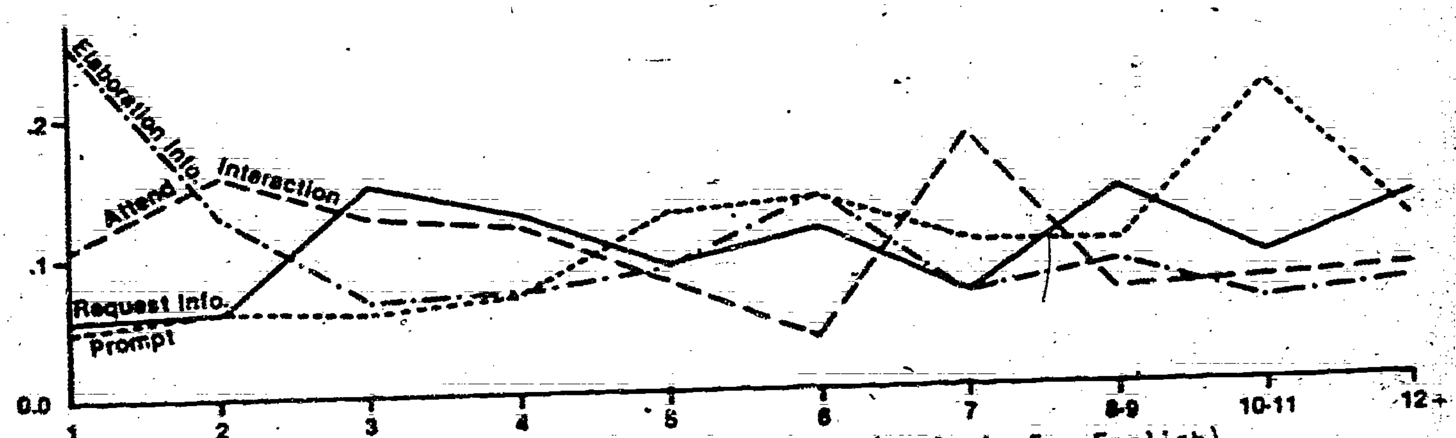

Figure 6.3. Function of previous turn (Analysis 5a, English).

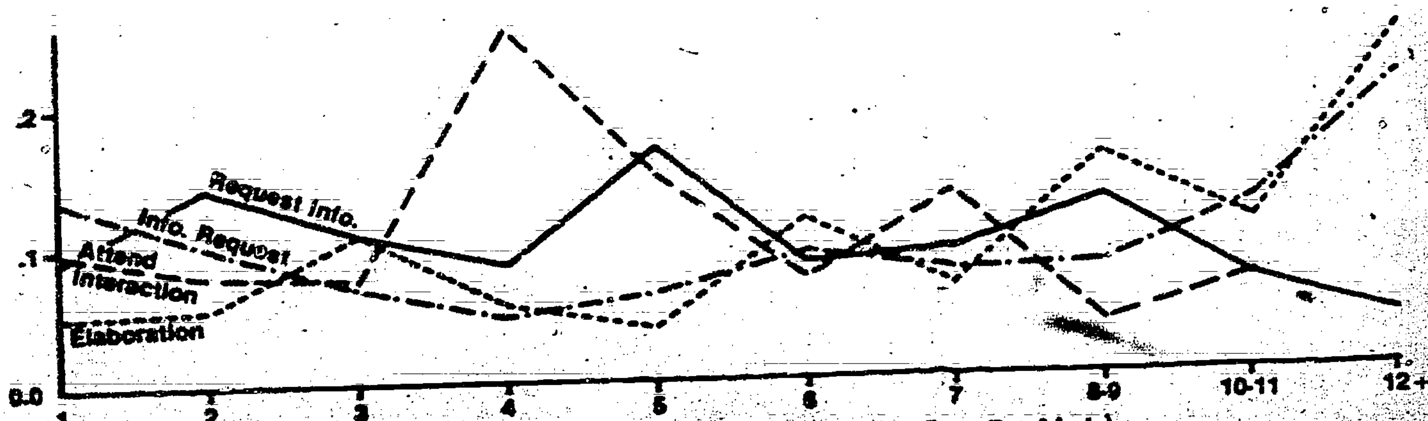

Figure 6.4. Function of turn (Analysis 5a, English).

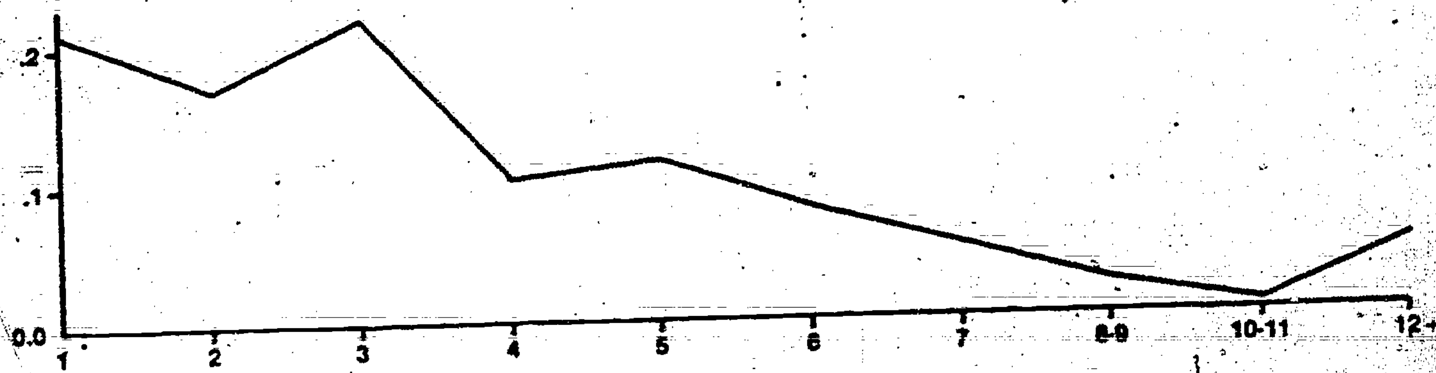

Figure 7.1. Overall effect (Analysis 5b, Spanish).

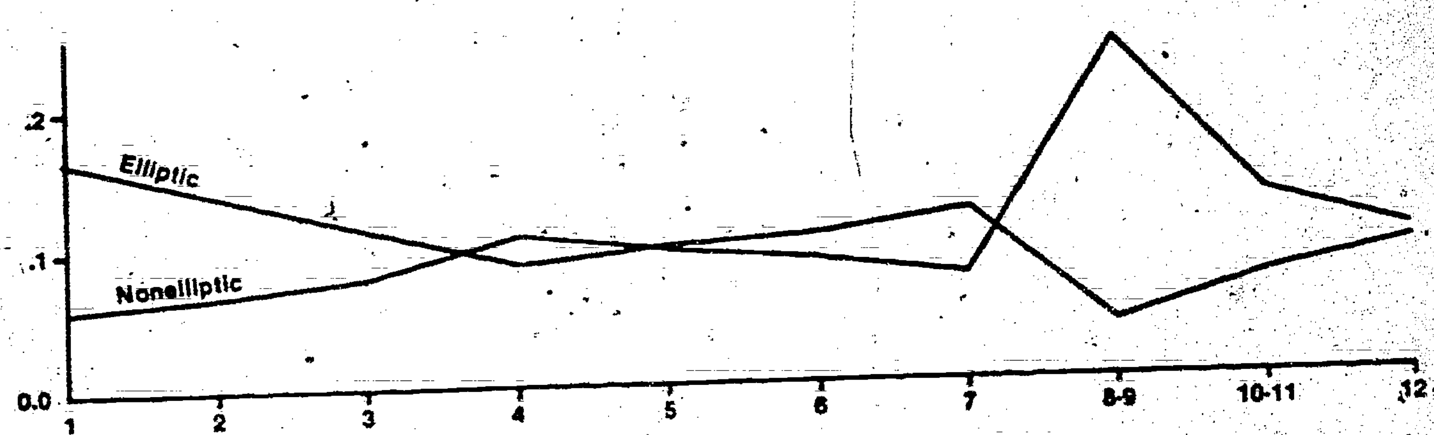

Figure 7.2. Ellipsis (Analysis 5b, Spanish).

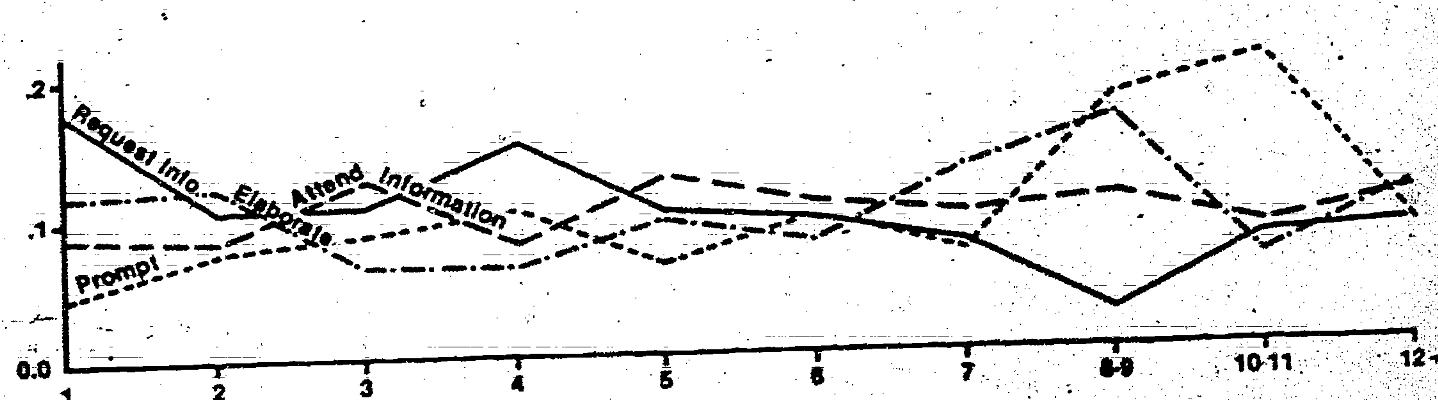

Figure 7.3. Function of previous turn (Analysis 5b, Spanish).

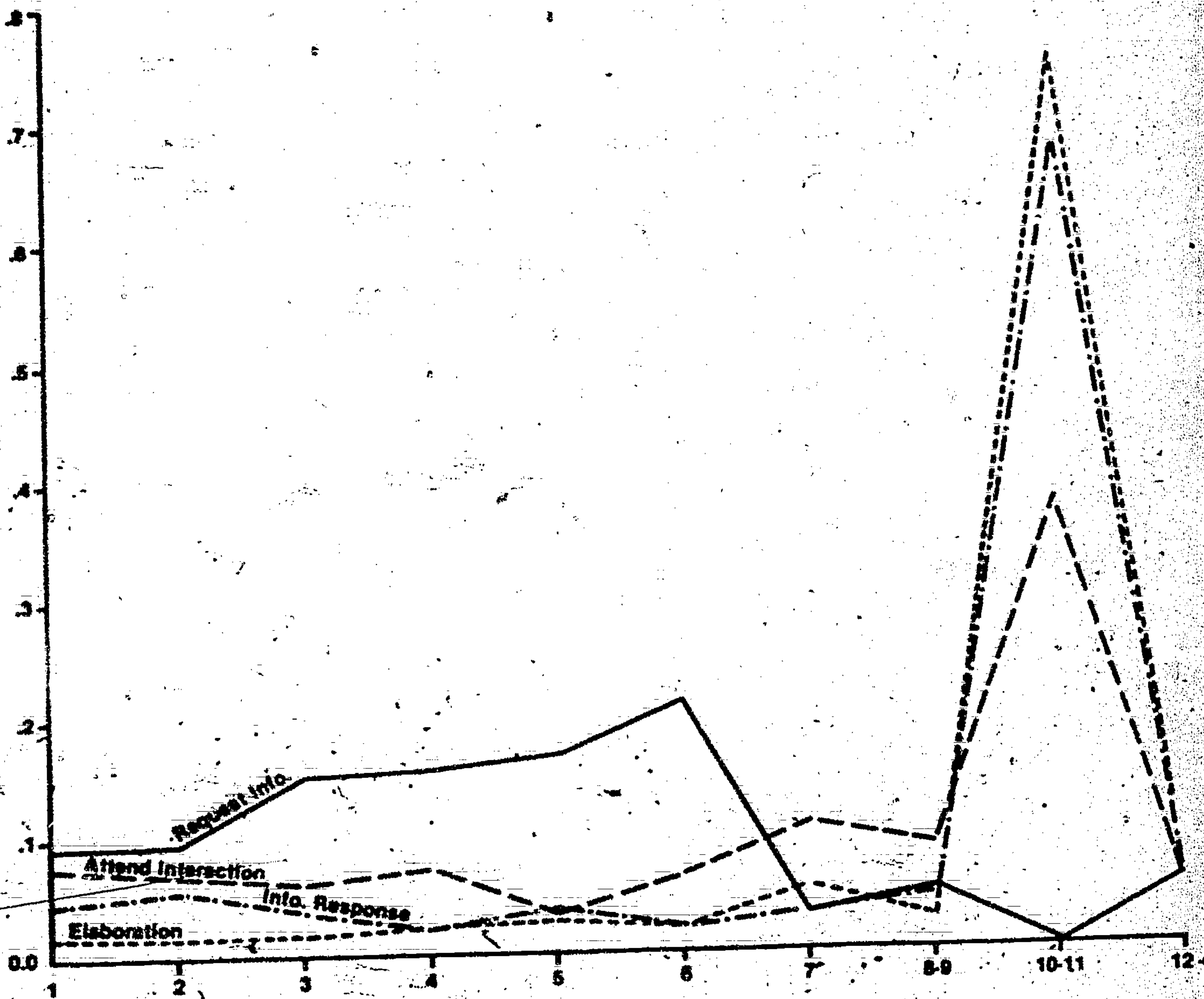

Figure 7.4. Function of turn (Analysis 5b, Spanish).

Ellipsis factors. The ellipsis factors show the expected effect that short utterances are more likely to be elliptic than non-elliptic, and that the opposite relationship holds at the high end of the length scale. There is essentially no effect in the mid range. The ellipsis effect is considerably mitigated by maintaining Number of Clauses as a factor group. This is likely due to the fact that elliptic utterances are almost always less than a full clause, and what shows as an ellipsis effect in Analyses 3 and 5 is in part a clause effect in Analysis 1. Splitting the data set by language shows the ellipsis effect to be greater in English than in Spanish.

Clause factors. The general effect of the Clause factors in Analysis 1 is somewhat of a mirror image of the Overall Effect. Partial Clause shows a positive effect at Length 4; 1 clause has positive effects as Lengths 2 and 4; and more than one clause is weighted to the Lengths 8 and over.

Discourse function. The effects related to function are rather more complex. The patterns across Analyses 1 and 3 are generally comparable in both shape and magnitude. Considering the two languages seperately, however, shows quite different effects across the two languages. Considering the Function of Previous Turn, Requests for Information and Prompts tend to be weighted against very short utterances, with Request for Information showing a positive effect in the middle lengths, and Prompts a positive effect for longer utterances. When the previous turn was not directly an invitation for the speaker to give additional information, that is, when the previous turn itself consisted chiefly of new information, or was directed to the nature of the interaction process, shorter utterances by the target speaker are weighted more heavily.

Somewhat opposite effects are shown for the function of the measured turn itself. Elaboration is generally weighted toward the greater lengths, and Information Response to the middle range. The pattern shown here for the Spanish only data in Analysis 5b is suspect, due to a very small number of observations in some categories at higher lengths.

Separating Out the Effects of Discourse Context

The findings presented here on the effects of discourse context on length of utterance strongly suggest that any attempt to relate utterance length to linguistic development of children of this age or older must either control for discourse context, or develop a system for accommodating the influences of discourse contexts which are external to the speaker. The analyses presented above provide two different bases by which the influences of discourse context can be partialled out. Using multiple regression, it is possible to calculate expected lengths and show how individual children differ from the aggregate. Using the maximum likelihood estimates, it is possible to calculate expected frequency distributions across lengths, and again show how individual children differ from those expected distributions.

We will consider first the multiple regression. Using the B coefficients from the final step of the regression we calculate the predicted length of utterance for each discourse context, defined across the five discourse variables. The difference between those predicted lengths, and the actual observation lengths for each utterance for each child are the <u>residuals</u>. These residuals are conventionally used to calculate the error of the regression. Here we interpret the residual as the child's contribution to utterance length, once the effects of the discourse contexts, pooled across all children, are subtracted. For each child we calculate the mean residual, and use that value as a score for the child. These scores are shown under the column labelled "Residuals" in Table 10.

From the maximum likelihood calculations we take a rather different approach to separating out the effects of discourse contexts. The maximum likelihood estimates provide probabilities that utterances of any particular length will be observed in each discourse context. Actually, we are not interested in the probability of occurrence of an utterance of a particular length, but in the probability that an utterance of <u>at least</u> that particular length will occur in a given discourse context. In this sense, length is now interpreted as an ordinal scale; lengths are ordered, but no assumptions are made related to interval between lengths. These probabilities may be readily estimated from the cumulative frequency distribution of utterances, with frequencies ranked by length of utterance. The cumulative distribution is shown for this data set in Figure 8. A set of weights may be derived from these probabilities by the simple calculation $w_i = 1 - p_i$, where p_i represents the probability of an utterance of at least length i, and w_i is the assigned weight. In an information theory sense, this may be interpreted as utterances of high probability yielding little information about the language development of a child, while utterances of low probability yield more information, and are weighted proportionately greater.

These probabilities form a scale rather different from a scale based on word count. The two scales are compared in Table 9. The percentage of change in the assigned value for each additional word in the utterance is for very short utterances, much greater for the weighted index than for the integral count. The percent of change is identical at six words and then tapers off very quickly for the weighted index. This has the effect of stretching the scale up to about the mean length of utterance for the children in this sample, and then compressing the upper reaches of the scale. Intuititively, this seems a desirable counter to the apparent bias introduced into small samples (e.g., the ten utterances of the BINL sampling procedure) by an occasional very long but atypical sentence.

This weighting procedure also offers a very straightforward way of incorporating the effects of discourse context. Weights may be

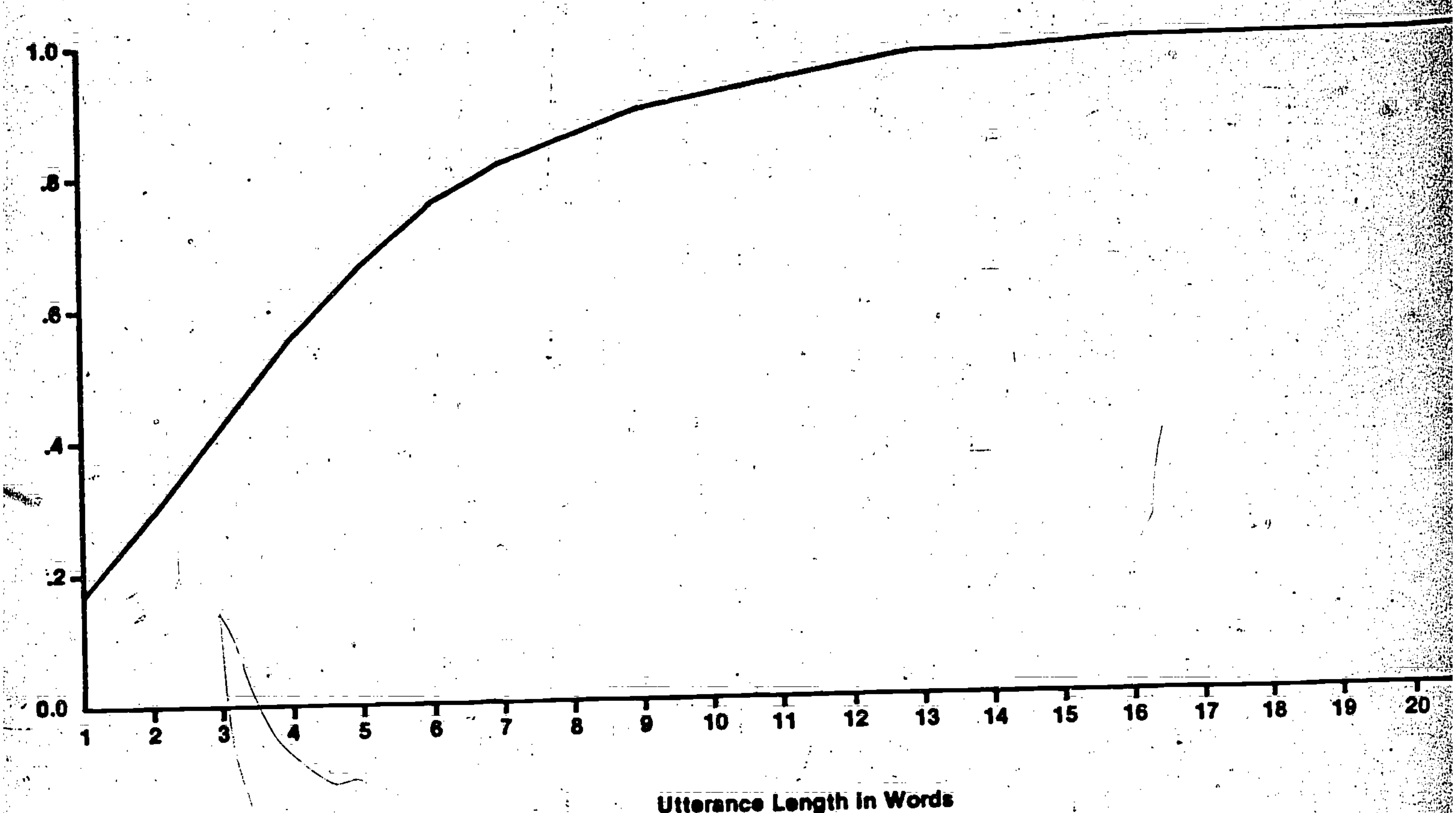

Figure 8. Cumulative frequency distribution.

Table 9

Length	% Increment	Weight (1-p_i)	% Increment	N
1	--	.00	--	219
2	1.0	.17	--	153
3	.50	.29	.71	176
4	.33	.42	.45	165
5	.25	.55	.31	139
6	.20	.66	.20	115
7	.16	.75	.14	85
8	.14	.81	.08	55
9	.13	.85	.05	40
10	.11	.89	.05	26
11	.10	.91	.02	35
12	.08	.93	.02	21
13	.08	.95	.02	24
14	.07	.97	.02	8
15	.07	.97	.01	12
16	.07	.98	.01	6
17	.06	.99	.00	4
18	.06	.99	.00	2
19	.06	.99	.00	3
20	.05	.99	.00	1
21	.05	.99	.00	1
22	.05	1.0	.00	1
23	.05	1.0	.00	1
24	.04	1.0	.00	1
25	.04	1.0	.00	0
26	.04	1.0	.00	1
27	.04	1.0	.00	0
28	.04	1.0	.00	0
29	.04	1.0	.00	0
30	.03	1.0	.00	0
31	.03	1.0	.00	1

calculated independently from the frequency distributions observed in each discourse context. Cumulative frequency distributions for elliptic and nonelliptic utterances in English and Spanish are graphed in Figure 9. From this graph it may be seen that, in both English and Spanish, elliptic utterances occur at higher relative frequencies in the shorter length categories than do nonelliptic utterances. Under the weighting schema suggested above, higher weights would be assigned to elliptic utterances than to nonelliptic utterances of the same length, with the differential decreasing as utterances get longer.

In principle, it would be possible to continue to subdivide the data set by the various discourse variables that have been considered here, and derive separate weights for each context. There are two problems with this, one practical, the other theoretical. The practical problem is again the distribution of observations across particular contexts, i.e., cells defined by each combination of independent variables, in the data set. The number of observations in many of the discourse contexts in this sample is so small that a great deal of sampling variation can be expected. The theoretical problem is that the procedure treats all contexts as independent and makes no assumptions that there are common factors operating in varying combinations across the contexts. The coefficients which derive from the maximum likelihood estimation provide a solution to both of these problems. The model assumes that a limited number of factors are independent, and that the factors rather than the many contexts which they jointly define, produce the observed differences in frequency distributions across the various discourse contexts. Under this assumption of independence, it is possible to use related contexts for the estimation of effects in contexts where there are very few observations.

To incorporate the discourse effects estimated by the maximum likelihood procedure, we use the same weighting procedure discussed above: One minus the estimated probability of an utterance at least as long as that observed in any particular discourse context.

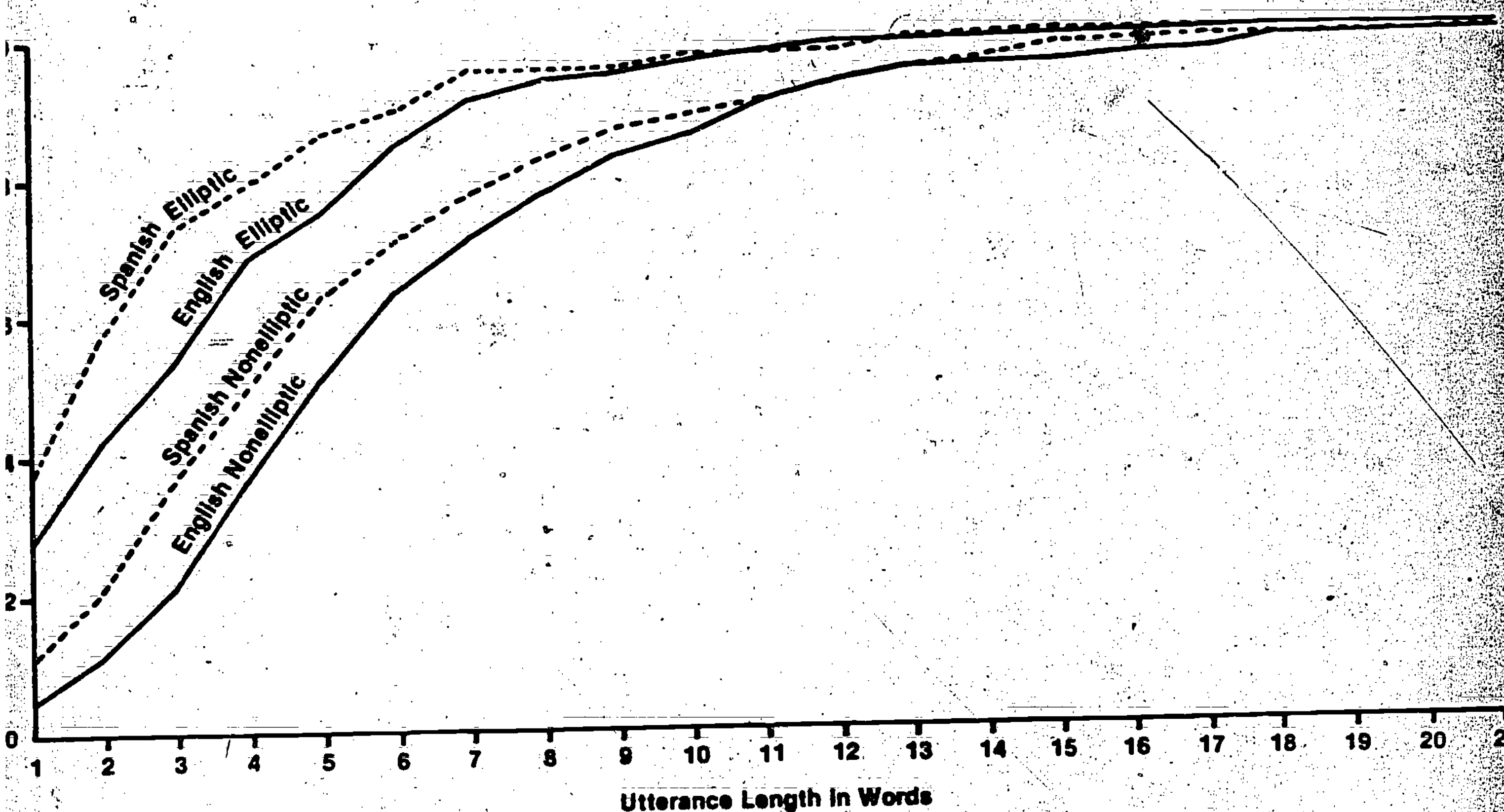

Figure 9. Cumulative frequency distributions, language x Ellipsis

Table 10

Individual Scores on Language Measures

		ENGLISH								SPANISH							
Child ID*	Age (Year; Month)	Judges' Profic. Rating	MLU	Residual	Mean Weights, Max Likelihood Analyses 1	2	3	4	5a	Judges' Profic. Rating	MLU	Residual	Mean Weights, Max Likelihood Analyses 1	2	3	4	5b
PA	9;8	23.5	8.1	2.86	.50	.58	.58	.59	.57	70.0	6.62	2.29	.45	.58	.59	.58	.57
RR	9;3	52.5	6.24	0.63	.46	.45	.47	.45	.46	29.0	5.50	0.82	.35	.45	.47	.46	.47
MF	8;11	52.5	5.94	0.76	.41	.49	.51	.48	.52	51.0	5.72	0.05	.41	.45	.47	.45	.48
MS	8;8	49.0	5.51	1.61	.32	.49	.45	.49	.45	51.0	3.54	0.90	.42	.54	.44	.55	.45
JM	8;7	31.0	5.96	0.92	.40	.44	.46	.45	.46	54.5	3.61	-0.53	.30	.35	.37	.35	.36
DF	7;6	38.5	6.32	1.14	.44	.52	.54	.53	.53	55.0	5.73	0.62	.43	.45	.48	.46	.48
SR#	7;5									48.5	5.30	0.20	.33	.39	.41	.39	.41
VM	7;5	39.5	5.86	0.69	.46	.48	.51	.49	.49	46.5	3.94	-0.68	.31	.36	.39	.36	.38
JT	7;2	32.0	4.50	-0.08	.33	.42	.47	.42	.45	53.0	10.54	4.69	.40	.71	.72	.72	.73
VS	7;2	23.0	3.12	-0.61	.30	.35	.29	.35	.29	28.0	2.79	-0.35	.27	.37	.30	.38	.30
ER	6;11	50.5	5.05	0.95	.46	.43	.45	.43	.45	20.5	4.84	0.52	.46	.50	.47	.50	.64
LA	6;9	20.0	2.78	-1.84	.29	.17	.17	.17	.18	64.0	5.38	0.92	.35	.48	.51	.48	.49
MR#	5;6									35.0	2.07	-2.33	.17	.14	.16	.14	.17
GH#	5;0									30.5	2.82	-1.15	.22	.26	.28	.25	.28
RT	4;8	17.5	5.38	-0.14	.47	.39	.42	.40	.41	32.0	9.21	3.72	.41	.64	.66	.65	.66
TN	3;8	15.5	2.48	-0.68	.18	.27	.20	.26	.20	11.5	1.91	-0.74	.18	.20	.13	.20	.13
Mean	7;5	34.2	5.17	0.49	.39	.42	.42	.42	.42	42.53	4.97	0.56	.34	.43	.43	.43	.44
S.D.	4;5	13.8	1.59	1.12	.09	.10	.12	.11	.12	16.28	2.39	1.72	.09	.15	.15	.15	.16

*From Garcia, et al. (1982)

#...d not respond to BINL in English; not included in English tabulations.

These weights are, thus, equivalent to the values of the predicted cumulative frequency distributions in Figures 1-3 where the weight for length i is equal to the cumulative probability for length i-1, for all i = 2, 3, . . . k. In all contexts, the weight for Length 1 is 0.0.

Comparison of Weightings

From each of the maximum likelihood estimates, weights based on the estimated cumulative probability were derived for each sentence. Mean weights were then calculated for each child in each language under each of the three analyses (Table 10). These mean weights provide a basis for at least an initial test of whether the general notion of weights based on frequency of occurrence and discourse context are of utility in estimating language development, and if so, which of the alternative approaches is preferable. The problem, of course, is the identification of an independent standard of language development, such that the proposed measure can be evaluated.

No satisfactory standard is available for this data set, and so two general indicators will be considered, both separately and jointly. These are age and judges' holistic ratings of language proficiency. A third indicator, scores on the BINL proficiency measure, i.e., their complexity scores, was not considered. The correlation of these complexity scores with MLU approaches unity (English, r = .998; Spanish, r = .975) and they cannot be treated as an independent indicator.

Age as an independent measure of language development. Age is an extremely precarious indicator of language development for a bilingual population. The children in this sample have highly diverse experiences in each of their languages. Some of the children are from homes where Spanish was the primary, perhaps only, language until they entered school. Others are from homes where English is the primary language for at least some dyads. Some of the children are in full bilingual programs in school; others are in regular programs, with or without extra instruction in English. Nonetheless, we expect that there will be a generally positive relationship between age and language development in both languages.

Holistic ratings of language proficiency. The holistic ratings of language proficiency were based on approximately ten pages of a transcript from each administration of the BINL, English and Spanish, for each sibling pair, as previously discussed in the method section. Five adult judges who are fluently bilingual in English and Spanish, and who had not had direct contact with the children in the study, were asked to rate each child's language (about 45 turns on the average) on a scale of one to ten, with the following instruction:

> We would like to get an idea of how well the children in our study speak English and Spanish. On the basis of your impressions from looking at the transcript provided, please rank the two children (whose initials or names appear above) on a scale of 1-10, with 10 being excellent, 5 being average, and 1 being poor. Use your own criteria as the basis for these evaluations/rankings.

To verify that judges were responding in reasonably comparable ways to the task, we converted the ratings to ranks, and calculated Kendall's Coefficient of Concordance (Siegel, 1956) for both the English and the Spanish samples. There was significant concurrence across judges for each sample (English: Kendall's $W = 0.531$, χ^2 $(12) = 31.87$, $p < .01$; Spanish: Kendall's $W = 0.488$, χ^2 $(15) = 36.66$, $p < .001$). The sum of ranks across the five judges for each child in each language is given in Table 10.

Correlations for age, mean judge's proficiency rating, MLU, and mean weights calculated under each of the three maximum likelihood procedures are given in Table 11a and 11b for English and Spanish, respectively.

Table 11a

Correlation Coefficients for Age and Eight Indicators of Language Proficiency

ENGLISH

	Age	Judges	MLU	Residual	Max Likelihood Estimates 1	2	3	4	5a
Age	--	.602	.693	.649	.495	.649	.638	.643	.661
Judges	*	--	.442	.451	.402	.537	.556	.510	.579
MLU	*	n.s.	--	.893	.853	.909	.930	.920	.936
Residual	*	n.s.	**	--	.629	.937	.864	.935	.872
M.L. 1	n.s.	n.s.	**	*	--	.707	.808	.728	.807
M.L. 2	*	n.s.	**	**	*	--	.962	.998	.964
M.L. 3	*	*	**	**	**	**	--	.966	.998
M.L. 4	*	n.s.	**	**	**	**	**	--	.966
M.L. 5a	*	*	**	**	**	**	**	**	--

Table 11b

SPANISH

	Age	Judges	MLU	Residual	Max Likelihood Estimates 1	2	3	4	5b
Age	--	.627	.195	.166	.579	.406	.406	.405	.398
Judges	**	--	.384	.319	.474	.464	.504	.448	.483
MLU	n.s.	n.s.	--	.934	.684	.886	.934	.882	.940
Residual	n.s.	n.s.	**	--	.687	.943	.909	.942	.913
M.L. 1	*	n.s.	**	**	--	.852	.859	.854	.859
M.L. 2	n.s.	n.s.	**	**	**	--	.965	.999	.967
M.L. 3	n.s.	*	**	**	**	**	--	.961	.998
M.L. 4	n.s.	n.s.	**	**	**	**	**	--	.965
M.L. 5b	n.s.	n.s.	**	**	**	**	**	**	--

* $p < .05$; ** $p < .01$

Patterns of relationships among the measures of language development.

The straight measure of length, MLU is significantly correlated with age only in English. MLU does not correlate significantly with the judges' rankings in either language. Thus MLU by itself does not seem to show a particularly consistent relationship to development. For both languages, the judges' rankings correlate with age at about r = .6.

The length measures which variously accomodate discourse context produce results which do not differ greatly from MLU; all but one correlate with MLU in the r =.85—.95 range. The one exception is Analysis 1 in Spanish, r = .687. The residuals from the multiple regression show relationships to age and judges' rankings that are comparable to, or somewhat lower than those for MLU. Again the correlation is significant only with age, and only in English.

The patterns for the maximum likelihood weights are somewhat mixed. All except Analysis 1 correlate significantly with age in English; exactly the opposite is true in Spanish. There only Analysis 1 correlates significantly with age. Analysis 3 correlates with the judges' rankings in both languages. Analysis 5a also correlates significantly with both age and judges' rankings in English, but not in Spanish.

As might be anticipated from comparison of the log likelihoods in Table 7, Analysis 2 and Analysis 4 produce essentially identical results. Also, Analysis 3 is extremely similar to Analyses 5a and 5b.

Recognizing that there are, in this sample of children, numerous situations which intervene in the expected relationship between age and language development, we compared the various indicators derived from length with age and the judges' proficiency ratings, considered jointly. Separate multiple regressions were computed for MLU, for Residuals, and for each of the weightings from the Maximum Likelihood Estimations in English and in Spanish, for Age and Judges' Ranks as independent variables. These are summarized in Table 12.

Table 12

Summary of Multiple Regressions of Length Measures on Age and Judges' Proficiency Rankings

ENGLISH

Dependent Measure	Multiple R	R2	Standard Error
MLU	*.694	.482	1.259
Residuals	.653	.427	0.970
M.L. 1	.512	.262	0.089
M.L. 2	*.674	.454	0.088
M.L. 3	*.673	.453	0.103
M.L. 4	.662	.438	0.092
M.L. 5a	*.699	.488	0.096

SPANISH

Dependent Measure	Multiple R	R2	Standard Error
MLU	.388	.151	2.364
Residuals	.321	.103	1.806
M.L. 1	.597	.356	0.081
M.L. 2	.486	.237	0.142
M.L. 3	.517	.268	0.148
M.L. 4	.476	.226	0.146
M.L. 5b	.498	.248	0.149

*p < .05

Considering Age and the judges' rankings jointly had essentially no effect whatsoever for MLU, or for the residuals from the regression. It did, however, generally increase the relationship shown to the various weightings, but only very slightly.

Conclusions

Our general contention that discourse function or context influences utterance length seems well confirmed in these language samples. Any attempt to draw inferences related to language development from examination of measures of length must accommodate these effects, at least for children as old as those observed here. The effect of discourse on length can be identified variously in terms of function of turn or function of previous turn, or syntactically in terms of ellipsis or number of clauses. When number of clauses is considered, the effect of ellipsis is largely obviated, but the effects of discourse function are largely unchanged. Noting the speaker of the previous turn does not seem to be preferable to noting the function of the previous turn. Nothing is gained by considering both speaker and function jointly.

The differences across languages seem fairly small. There was no language effect in the analyses of variance. Under maximum likelihood, language factors were also fairly small. When the data set was partitioned by language, the factors for function show rather different patterns. These differences, however, had essentially no effect on the derived weightings: Analysis 3 and Analysis 5a,b produced almost identical results.

From the maximum likelihood calculations, it is possible to derive predicted probability distributions for length. These weights can, in turn, replace length as a language development indicator. The resulting weights correlate more highly with judges' holistic ratings of proficiency than does MLU, in both English and Spanish. In Spanish, the correlation of the weights to Age is substantially higher than the correlation of MLU; in English, they are roughly comparable, with consideration of number of clauses producing a somewhat lower correlation. A similar pattern results when Age and Judges' Rankings are considered jointly in multiple regression.

All of this is based on a rather sparse data set from relatively few children. Nonetheless, the concept of weights derived from frequency distributions of length defined across discourse contexts appears extremely promising. The longitudinal study from which these data derive allows the possibility of enlarging the language samples from each child, and for looking at children across time. The ability to measure change within children across time will provide a truer test of the utility of the procedures. Whether or not it will be possible to demonstrate all five of the desirable measurement characteristics suggested above remains to be determined.

References

Berdan, R. On the nature of linguistic variation. Unpublished doctoral dissertation, University of Texas at Austin, 1975.

Berdan, R., & Garcia, M. Measuring language development in bilingual children. Presentation at National Association for Bilingual Education Annual Conference, Detroit, MI, April 1982.

Berko-Gleason, J. The child's learning of English morphology. In A. Bar-Adon & W. F. Leopold (Eds.), Child Language. Englewood Cliffs, NJ: Prentice Hall, 1971. (Reprinted from Word, 1958, 14, 150-177.)

Bloom, L. Form and function in emerging grammars. Cambridge, MA: MIT Press, 1970.

Bloom, L., Rocissano, L., & Hood, L. Adult-child discourse: Developmental interaction between information processing and linguistic knowledge. Cognitive Psychology, 1976, 8(4), 521-552.

Braine, M. D. S. Length constraints, reduction rules, and holophrastic process in children's word combinations. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 448-456.

Brisk, M. The Spanish syntax of the pre-school Spanish-American: The case of New Mexican five-year-old children. Unpublished doctoral dissertation, University of New Mexico, 1972.

Brown, R. A first language. The early stages. Cambridge, MA: Harvard University Press, 1973.

Cazden, C. The acquisition of noun and verb inflections. Child Development, 1968, 39, 433-438.

Cedergren, H. J., & Sankoff, D. Variable rules: Performance as a statistical reflection of competence. Language, 1974, 50, 333-355.

Cohen, J. & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum, 1975.

Cox, D. R. The analysis of binary data. London: Methuen, 1970.

Dore, J. Conversational acts and the acquisition of language. In E. Ochs & B. B. Schieffelin (Eds.), Developmental Pragmatics. New York, NY: Academic Press, 1979.

Edwards, A. W. F. Likelihood. Cambridge: Cambridge University Press, 1972.

Garcia, M. Pa(ra) in locative phrases in the Spanish of El Paso-Juarez. Unpublished doctoral dissertation, Georgetown University, 1981.

Garcia, M., Veyna-Lopez, A., Siguenza, C., & Torres, M. Baseline data report of the Spanish-English longitudinal study. Working Paper 82-3W. Los Alamitos, CA: National Center for Bilingual Research, 1982.

Halliday, M., & Hasan, R. Cohesion in English. London: Longman, 1976.

Herbert, C. Basic inventory of natural language. Instructions manual (Revised ed.). San Bernardino, CA: Checpoint Systems, Inc., 1979.

Hofacker, C. A look at software for dichotomous dependent variables. In Perspective, 1982, 6(3), 4-7.

Hunt, K. W. Grammatical structures written at three grade levels. Research Report 3. Urbana, IL: National Council of Teachers of English, 1965.

Jones, R. H. Probability estimation using a multinomial logistic function. Journal of Statistical Computation and Simulation, 1975, 3, 315-329.

Labov, W. The social stratification of English in New York City. Washington, DC: Center for Applied Linguistics, 1966.

Lindsey, J. K. Likelihood analyses and tests for binary data. Applied Statistics, 1975, 24, 1-16.

Loban, W. Language development: Kindergarten through grade twelve. Research Report 18. Urbana, IL: National Council of Teachers of English, 1976.

Longhurst, T. M. (Ed.) Linguistic analysis of children's speech. New York: MCS Information Corporation, 1974.

Menyuk, P. Syntactic structures in the language development of children. In A. Bar-Adon and W. F. Leopold (Eds.), Child Language. Englewood Cliffs, NJ: Prentice Hall, 1971. (Reprinted from Child Development, 1963, 34, 407-422.

O'Donnell, R. C., Griffin, W. J., & Norris, R. C. Syntax of kindergarten and elementary school children: A transformational analysis. Research Report 8. Urbana, IL: National Council of Teachers of English, 1967.

Padilla, A., & Liebman, E. Language acquisition and the bilingual child. In J. Fishman & G. D. Keller (Eds.), Bilingual Education for Hispanic Students. New York, NY: Teacher's College Press, 1982.

Peters, A., Ostman, J., Larsen, T., & O'Connor, M. Manual for computer coding of bilingual data. Unpublished manuscript. Berkeley, CA: School of Education, University of California at Berkeley, 1982.

Peronard, M. Adquisición del lenguaje en un niño atípico: Retardo o desviación. Boletín de filosofía, 1977, 28, 139-152.

Rousseau, P., & Sankoff, D. Advances in variable rule methodology. In D. Sankoff (Ed.), Linguistic Variation. Models and Methods. New York, NY: Academic Press, 1978.

Sankoff, D. VARBRUL VERSION 2. Unpublished manuscript. Quebec, Canada: Centre de recherches mathematiques, 1975.

Siegel, S. Nonparemetric statistics for the behavioral sciences. New York, NY: McGraw-Hill Book Company, Inc., 1956.

Sinclair, J., & Coulthard, R. M. Towards an analysis of discourse. The English used by teachers and pupils. London: Oxford University Press, 1975.

Wells, G. Learning through interaction: The study of language development. Cambridge, MA: Oxford University Press, 1981.